电信技术普及丛书

汉字终端技术入门

周关兴 编著

人民邮电出版社

20207
TP3
155

电信技术普及丛书

汉字终端技术入门

周关兴 编著

Retro Workshop
果粉工作室

人民邮电出版社

内 容 提 要

本书为“电信技术普及丛书”之一。

汉字信息处理是一门正在蓬勃发展的新兴学科，汉字终端技术是汉字信息处理技术的重要组成部分。本书深入浅出地叙述了汉字终端技术的发展概况、汉字的基本特点、汉字输入和输出技术、汉字编码和通信技术、汉字终端的软件技术等方面的基本知识。特别对具有较高实用价值的各种汉字输入编码方案和各种最新的标准资料，都作了简要的介绍，对使用、维修、生产汉字终端设备的工作者具有较大的参考价值，可作为自学和培训的参考读物；还可以作为广大读者了解汉字信息处理技术特别是汉字终端技术的入门向导。

电信技术普及丛书

汉字终端技术入门

周关兴 编著

责任编辑：蔡效平

人民邮电出版社出版

北京东长安街27号

天津新华印刷一厂印刷

新华书店北京发行所发行

各地新华书店经售

开本：787×1092 1/32　　1990年2月　　第一版

印张：13 28/32　页数：222　1990年2月 天津第1次印刷

字数：319千字　　印数：1—6 500册

ISBN7-115-04161-X/TN·325

定价：5.40元

出版者的话

为了普及电信技术知识，特别是电信新技术知识，为我国的通信现代化服务，我们组织编写了一套“电信技术普及丛书”陆续出版。这套丛书的主要读者对象是具有中学以上文化水平、有一些电信基本知识的工人、管理干部和有关技术人员。在编写中，力求做到内容正确，概念清楚，深入浅出，通俗易懂；使读者读过一书后，能对某项技术的基本原理和主要情况有一个概括的了解，作为进一步学习的入门向导。我们殷切希望广大读者对这套丛书提出意见和建议，帮助我们做好这一工作。

Retro Workshop
果粉工作室

前　言

现代科技的迅速发展，使汉字信息处理技术的难题得到了解决。各种类型的汉字终端设备，包括汉字电传机、中西文打字机和文字处理机如雨后春笋般地不断涌现，并已在计算机、通信和办公自动化等领域得到了广泛的应用，成为促进我国四化建设的有力工具。为了适应广大读者希望了解和掌握汉字终端技术的要求，特编写了这本小册子。

本书力求通俗易懂、深入浅出地叙述汉字终端技术的发展概况、汉字的基本特点、汉字输入和输出技术、汉字编码和通信技术、汉字终端的软件技术等方面基本知识。特别对具有较高实用价值的各种汉字输入编码方案和各种最新的标准资料都作了简要的介绍。这对使用、维修和生产汉字终端设备的工作者，特别是从事电报和数据通信工作的技术人员，都具有较大的参考价值，可作为自学和培训的参考读物；还可作为广大读者了解汉字信息处理技术特别是汉字终端技术的入门向导。

在本书编写过程中，得到总参第六十一研究所领导及同志们的大力支持；并承中国中文信息学会常务理事、燕山信息部总工程师石云程同志审阅全书，提供了许多宝贵资料和修改意见。在此，谨向他们表示最衷心的感谢。

汉字信息处理是一门正在蓬勃发展的新兴学科。汉字终端技术是汉字信息处理技术的重要组成部分。作者希望本书的出版能对在我国普及汉字信息处理技术知识作些微薄贡献。由于本人水平和时间等条件所限，书中难免有遗漏或不当之处，敬请读者不吝指正。

作者　1989年3月于北京

果粉藏書

目　录

一、概述 …… (1)

1·1　什么是汉字终端 …… (1)
1·2　汉字终端系统的组成 …… (3)
1·3　汉字终端的代码流程 …… (8)
1·4　汉字终端技术的历史和现状 …… (11)
1·5　汉字终端技术的基本内容 …… (14)

二、漫话汉字 …… (16)

2·1　汉字的历史演变及特点 …… (16)
2·2　汉字的数量及使用频度 …… (20)
2·3　汉字的字形特性 …… (22)
2·4　汉字的字音特性 …… (35)

三、汉字输入技术 …… (39)

3·1　汉字输入技术综述 …… (39)
3·2　汉字输入编码 …… (43)
1.汉字输入编码概述 …… (43)
2.汉字输入编码的种类和特点 …… (50)
3·3　汉字输入编码方案实例介绍 …… (55)
1.基本笔画笔顺编码法 …… (55)
2.五笔画汉字输入法 …… (58)
3.汉字笔形编码法 …… (59)
4.五笔字型汉字输入法 …… (62)

5.五笔桥汉字输入法 ……(67)
6.联合45-3汉字拼形输入法 ……(69)
7.HPX汉字拼形输入法……(73)
8.仓颉字母输入法 ……(77)
9.大众汉字拼形输入技术 ……(83)
10.见字识码输入法……(85)
11.HYX汉字音形输入法 ……(89)
12.通信拼音字……(91)
13.双拼双音字词编码法……(94)
14.人大码汉字词语输入法……(97)
3·4 汉字输入键盘 ……(102)
1.标准键盘（小键盘） ……(103)
2.整字键盘（大键盘） ……(125)
3.专用键盘（中键盘） ……(142)
3·5 汉字识别 ……(144)
1.汉字识别技术概述 ……(145)
2.印刷体汉字识别 ……(148)
3.手写体汉字实时识别 ……(163)

四、汉字存储及输出技术 ……(169)

4·1 存储器概述 ……(170)
1.存储器的分类及特点 ……(170)
2.半导体存储器的基本工作过程 ……(176)
3.磁盘存储器的基本工作过程 ……(177)
4·2 汉字字形发生器 ……(180)
1.字形发生器的分类及特点 ……(181)
2.数字式字形发生器的原理 ……(185)
3.数字式字形发生器的应用 ……(194)
4·3 汉字显示器 ……(219)

1.概述 ……(219)
2.阴极射线显示管的基本工作原理 ……(225)
3.平板型显示器的基本工作原理 ……(231)
4.汉字显示器的基本构成和刷新方式 ……(240)
4·4 汉字打印机 ……(244)
1.打印机的分类和特点 ……(245)
2.点阵式汉字打印机的使用要求和主要性能 ……(248)
3.汉字打印机的基本原理及构成 ……(257)
4.常用点阵打印头的工作原理 ……(266)

五、汉字代码及汉字通信 ……(274)

5·1 汉字内部码和传输交换码 ……(274)
1.西文系统的内部码 ……(275)
2.七位标准代码表 ……(279)
3.汉字交换码 ……(281)
4.汉字内部码的设计要求 ……(288)
5.汉字内部码方案介绍 ……(292)

5·2 汉字控制功能码 ……(303)
1.七位标准代码表中的控制功能字符——C。集内容介绍 ……(303)
2.控制功能代码的扩充——国标GB2311—80内容简介 ……(313)
3.文字和字符成形设备的增补控制功能——国标GB5261—85内容简介 ……(321)

5·3 汉字通信 ……(335)
1.汉字数据通信概述 ……(336)
2.汉字数据通信的接口和规程 ……(358)

六、汉字终端的软件技术 ……(377)

6·1 微型计算机系统的组成 ……(377)

1.概述 ……(377)
2.微型计算机的硬件 ……(379)
3.微型计算机的软件 ……(389)
6.2 汇编语言及程序设计初步 ……(393)
1.概述 ……(393)
2.指令结构及其功能 ……(394)
3.汇编语言 ……(398)
4.程序设计初步 ……(404)
6.3 汉字终端的应用软件 ……(410)
1.键盘输入管理程序 ……(410)
2.汉字输出程序 ……(419)
3.汉字文本编辑程序 ……(427)

主要参考资料 ……(433)

Retro Workshop
果粉工作室

一、概　述

1·1　什么是汉字终端

终端，通常是指直接面向使用者或操作人员的设备。它能将人们提供的文字、声音、图象、数据等信息转变为电信号，输入到计算机或通信系统中进行处理或传输；并将计算机或通信系统输出的电信号变换成相应的文字、声音、图象或数据提供给使用者。

在计算机系统中，终端设备的作用是向计算机输入程序和数据，并接收计算机输出的处理结果。计算机终端的种类很多，大体上可分为远程批处理终端和交互式终端两大类。前者用于对远距离的主计算机进行批量数据输入输出作业，最简单的远程终端设备是由卡片阅读机、打印机及其通信控制器组成；后者又称会话式终端，是人与计算机系统之间以人——机交互方式（即会话方式）进行数据处理作业的工具，常用的有电传机、键盘显示器等。

在通信系统中，终端是开始进入或最终离开通信网的设备。它们都处在通信过程的最终点，通常由用户直接掌握使用，因此又常称为用户终端设备或末端设备。常用的有各种电话机、电传机、传真机、数据终端机等。

汉字终端是用于计算机或通信系统中进行汉字信息的输入输出及加工处理的一种终端设备，是使用汉字的人们实现文字信息处理和通信的直接工具。它一般应具有以下功能：

1. 汉字输入输出功能

使用者通过汉字终端向计算机输入汉字信息，计算机系统通过汉字终端输出汉字。也就是说，汉字终端的输入输出功能，是把“有形的汉字”转化为“无形的汉字”（即汉字代码）；又把“无形的汉字”（汉字代码）转化为“有形的汉字”。应当指出，由于汉字的字种繁多，字形结构复杂，汉字终端与通用的字母数字终端（即西文终端或字符终端）的文字输入输出功能相比，在技术复杂性上有很大的差异。此外，汉字终端除了能输入输出汉字外，还应具有字母数字的输入输出功能。

2. 通信功能和汉字信息处理功能

作为通信系统或计算机系统的输入输出设备，汉字终端与其它终端一样，都具有与系统（主机）进行通信的功能，它们之间必然有一个通信接口。汉字终端将汉字信息输入主机以后，按转换规则变成通用的字符信息组合（电码），送到主机进行加工处理或送到对方终端以实现汉字通信。单纯实现这类功能的汉字终端设备有汉字电传机、简易型汉字显示终端等。还有一类独立型的汉字终端设备，它除了完成汉字输入输出功能和通信功能外，还能实现汉字信息的加工处理，如汉字信息的分类、比较、检索、存储和交换等功能。这类设备的典型代表是具有汉字功能的微型计算机。

这里提一下汉字终端与汉字微型计算机的区别。尽管它们中的某些类型在硬件结构上没有多大的不同，但在软件结构和处理能力上是有差别的。一般说来，汉字微型计算机应具有完整的、属于本身的操作系统和各种系统软件，具有独立的信息

处理能力；而汉字终端的全部软件都是为实现汉字输入输出和汉字通信、人机对话等各种具体功能设计的。它通常不具备很强的信息处理能力，并不一定要提供传统的高级语言处理程序，而是使用简明的交互方式指示用户进行操作使用。上面提及的一些功能较强的汉字终端，同时具备了汉字微型机的功能，可以看作是一台汉字终端和一台汉字微型计算机的结合体。

1·2 汉字终端系统的组成

一台汉字终端，不论是手提式的中西文电子打字机，或是作为大、中型计算机系统的远程用户终端，其基本系统的组成和工作原理都大体相同。它们都是以微处理器为核心的一个专用型计算机系统，通常包括微处理器、随机存取存储器、只读存储器、汉字字形发生器（即汉字字模库）、标准字母数字键盘、显示器以及通信接口等工作部件。配置较全的汉字终端还包括汉字打印机、汉字整字键盘、小型磁盘驱动器等工作部件。它们的组成框图如图1·1所示。

图1·1中，实线部分是基本型（或称简易型）汉字终端系统组成部分；虚线部分是基本型汉字终端系统的扩充部分。当然，基本部分和扩充部分的区别不是绝对的。在一些需要文件输出的使用场合（如汉字电传机、中文打字机等），汉字印字机就成为不可缺少的基本组成部分了。

下面简要介绍一下各个组成部分的主要功能。

1. 微处理器

微处理器（μP）又称中央处理器（CPU），是汉字终端

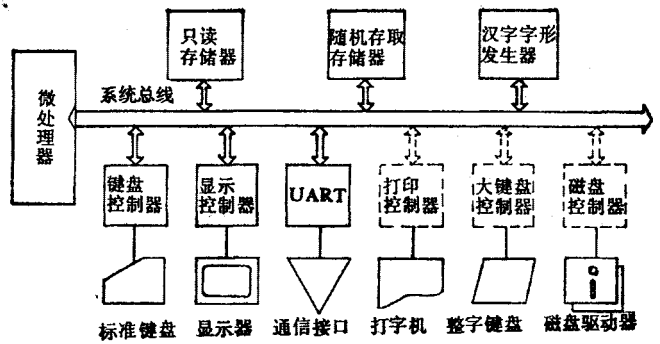


图 1·1 汉字终端系统的组成框图

的控制部件。它可以解释和执行指令，与外部交换信息，实现汉字处理过程中的各种运算和控制功能。

微处理器通常以同时处理的数据宽度来划分，有 1 位、4 位、8 位、16位和32位微处理器。汉字终端常用的是 8 位或16位微处理器。

2. 只读存储器和随机存取存储器

只读存储器（ROM）和随机存取存储器（PAM）统称为主存储器，又称半导体存储器，是汉字终端用来存放程序和数据的部件。只读存储器通常用来存放程序和固定不变的数据，以供CPU执行各种控制和处理功能时使用。随机存取存储器用来存放运算过程中的中间结果，以及作为输入输出数据的缓冲存储器。

半导体存储器的存储容量通常以字节为单位，汉字终端一般需要十几K到几十K字节（1 K = 1024字节）。

3. 系统总线

系统总线是微处理器与主存储器、输入/输出接口电路(即图中的各种控制器）之间传送信息的通道，它由三个部分组成：

（1）数据总线，用于传送数据和指令，可以来回两个方向传送。数据总线的宽度决定了一次操作中能传送的二进制位数的最大值，一般与微处理器处理数据的字长相同。同时，它也是确定微处理器和微型计算机“位数”的依据。8位微处理器的数据总线的宽度为8位。

（2）地址总线，用于传送数据和指令的存储地址。地址总线上的数据（即地址值）由微处理器给出后，按一个方向传送到存储器或输入/输出接口电路。地址总线的宽度决定了微处理器能够寻址的最大范围，8位微处理器通常有16条地址线，可以寻址的最大范围为 $2^{16}=65536$个地址单元。

（3）控制总线，用于传送各种控制信号，管理数据总线和地址总线上的信息传送。控制总线通常有十几条控制线，可分为两类：一类是微处理器向外界发出的控制信号，如读/写、选通等；另一类是外界向微处理器送入的控制和状态信号，如中断请求、暂停等。

4. 汉字字形发生器

汉字字形发生器又称为汉字字模库或字库，是汉字终端和汉字计算机系统所特有的重要组成部件。它通常是由一块或几块超大规模集成电路做成的只读存储器，专门用来存放汉字字形点阵（或特征）数据。

由于汉字字数多，字形点阵数据量很大。通常用来存储

7500个汉字和字符的简易型字形发生器就需要240K字节的存储容量。在大规模集成电路技术没有开发应用的时候，汉字字形发生器是汉字终端设备的一大“包袱”，往往需要做成单独一个机箱甚至机柜。现在，只需一块姆指般大小的存储器芯片，就可以存放国标GB2312-80收集的全部汉字和字符图形数据了。这就为汉字终端的应用和发展创造了极为有利的条件。

5. 汉字输入键盘

汉字输入键盘是汉字终端接收外界输入信息的主要设备。通常配置与西文打字机、计算机、电传机等设备相兼容的标准字母数字键盘，简称标准键盘或小键盘。用标准键盘输入汉字信息，需要经过汉字编码，即汉字输入编码。汉字输入编码目前有多种方案，大体上可分为形码、音码和音形码几类。使用汉字输入编码，一般需要经过一段时间的学习和记忆。为了方便不熟悉或不了解汉字输入编码的人员使用汉字终端，也可以配置汉字整字键盘（或称为大键盘）作为汉字的输入设备。

汉字输入键盘是一种外部设备，需经过键盘控制器与微处理器的系统总线连接。

6. 汉字显示器

汉字显示器是汉字终端用来输出汉字信息的主要设备之一，同时可以作为汉字输入时的监视设备和文字编辑显示设备。汉字显示器通常采用阴极射线管（CRT）显示器；在便携式的汉字终端设备中，也可采用平板型的显示器（如液晶显示器、电致发光或等离子体显示器）来显示汉字信息。

由于显示器的响应速度快，没有噪音及机械磨损部件，制造成本较低，因此在汉字终端上得到了广泛的应用。

7. 通信接口

通信接口是汉字终端与主计算机或其它通信设备交换信息的通道，通常采用半双工或全双工的工作方式，可以交替或同时地双向串行传送信息。通信接口通常采用统一的标准，通过通用同步或异步收发器（即通信接口控制器）与系统总线相连接，以实现收发信号的串/并变换，并选择适当的通信速率。

8. 汉字印字机

汉字印字机是实现汉字文件打印输出的主要设备。使用最广的汉字印字机是针式（点阵）打印机。有16针、24针和32针的针式打印机。简易型的终端设备还选用9针的西文打印机，使用“拼接”的方法来打印汉字，以求减少费用开支。除了针式打印机外，各种新型的印字机正在不断开发问世，如喷墨式印字机，热敏式和热转印式印字机以及液晶式和激光式印字机等，开始得到越来越多的应用。

针式打印机通过打印控制器与系统总线相连接，根据微处理器提供的数据和控制信号印字输出。在整机结构方面，打印机与主机的配置可以合为一体；但通常都是把二者分开配置，用短电缆相互连接。

9. 磁盘驱动器

磁盘驱动器是汉字终端的外存储器，或称辅助存储器。它是用来存放大量不参加当前运算处理的文件和数据。外存储器的存储容量比内存储器大得多，但存取信息的速度较慢。

汉字终端常用的磁盘驱动器有 $5\frac{1}{4}$ 英寸和 $3\frac{1}{2}$ 英寸的小型

软磁盘驱动器或硬盘驱动器。存储容量由几百K字节到几十兆字节。近来市场上已出现了激光盘存储器，每张小盘的存储容量达600兆字节。在简易型的汉字终端设备中，为了降低成本也有用普通的盒式磁带录音机作为外存储器。

10. 汉字整字键盘及其它输入设备

汉字整字键盘又称大键盘，是在类似机械式中文打字机键盘的平面上，按一定规则排列着几千个常用汉字。使用者用手指按压或用一支特殊的笔触点以后，就可将被选中的汉字输入到机器中去。对于在盘面上没有收集的汉字，则需用拼字或其它盘外字处理办法输入。

由于不论小键盘输入或大键盘输入都有一定的局限性，近年来，手写体汉字识别输入装置、印刷体汉字输入装置和汉语语音识别装置等各种新的汉字输入设备正在不断开发并逐步投入应用。

1·3 汉字终端的代码流程

从信息交换的角度看，汉字终端应具有两种基本功能：第一种是人与终端之间的信息交换，第二种是终端与主机（或另一个终端）之间的信息交换。

实现第一种信息交换，就是要求把人们所熟悉的信息形式（如文字、图象、声音）转换成机器能辨认的符号（即代码）；实现第二种信息交换，就是要求在机器与机器之间规定统一的符号代码及传输规程。这两个基本要求，不论对汉字终端或者西文终端来讲是共同的。

众所周知，西文拼音文字借助拼音规则，可以只用少量的

字母，就能拼写成表达各种“音”和“意”的文字。因此，人们使用西文终端输入拼音文字时，就不需用“输入码”，只要顺序按键输入字母就行了。由于字母的字形结构比较简单，通常只需 7 × 9 点阵数据便能表示。加之在设计其内部码和传输交换码时已做到了尽可能地相互一致，因而西文终端的代码转换影射过程也就变得非常简单。当按键输入一个字母，机器将它变换成相应的代码以后，在内部处理和传输交换时，仍然是这个代码。在某些计算机系统中（如IBM系统），为了处理上的方便，以某一种代码（如EBCDIC码）作为本系统机器的内部码，当该系统要与其它系统交换信息时，则必须先将它的内部码变换成符合统一标准的信息交换码（如ASCII码）。但对汉字来说，由于它的特殊性，汉字终端在实现上述两种信息交换中的代码转换过程，要比西文终端复杂得多。

现在，我们以图1·2来说明汉字终端的代码转换流程。操作者通过键盘将汉字信息以“输入码”的数据形式送入主控制

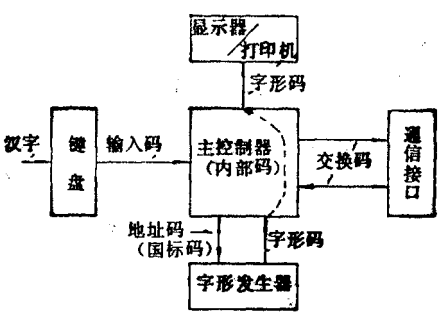


图 1·2　汉字终端代码转换流程示意图

器。机器接收后通过查表或运算操作，将输入码转换成便于机器处理的“内部码”，然后在内存储器的指定地址单元存储起来。当需要输出一个汉字时，主控制器应向字形发生器送入这

个汉字的“地址码”，以读出该字的字形数据(即“字形码”)，并将字形码送往显示器或打印机的缓冲存储器。然后通过执行显示或打印输出程序，就可在屏幕或打印纸上看到机器输出的字形。当需要将汉字信息送往主机或其它终端时，主控制器执行相应的通信程序，把汉字内部码转换成统一的传输交换码，并将它发送到通信线路上去。显然，上述的代码频繁转换过程是汉字终端设备所特有的。

那末，能不能把汉字终端的代码转换过程简化呢？我们可以用电报码制为基础的汉字电传机为例，来说明这个问题。

电报码是进行汉字电报通信的一种标准代码，它用 0 ～ 9 十位数中的四位数码作为一个汉字的码组。当使用西文电传机传送汉字信息时，发送方的操作员必需先将汉字翻译成电报码才能输入机器；接收方将收到的电报码再译成汉字。现在，有的汉字电传机仍沿用这种制式。汉字输入码、内部码和传输交换码都采用统一的电报码。这样，虽然避免了频繁的代码转换过程，但是在使用上和设计上是不方便和不经济的。因为，采用电报码作为汉字输入码，四个数字表示一个汉字，基本上无规律可循，全凭硬记。这对大多数非邮电通信报务专业的使用者来讲，显然是很难做到的。这也是我国的汉字用户电报通信难以普遍应用的主要原因之一。如果直接采用电报码作为机器内部码，则在计算机资源的利用上是不经济的。电报码的四位数字，如果把它们作为四个字符处理，则每个汉字就要用四个字节的内部码表示，会增加机器的软硬件开销；如果把它们作为十六进制数（或十进制数）处理，四位数可并为两个字节。但由于与国标计算机的汉字代码不一致，还需再进行一次代码转换，才能利用现有的计算机软硬件资源。

由此可见，为了使汉字终端使用方便，以及保持机器设计

的经济性和合理性，汉字的输入码、内部码和传输交换码等应有所区别；同时要求它们之间能比较方便地进行相互转换。

从代码的角度看，一台汉字终端就是一个各种汉字代码的转换系统。只要掌握住各种汉字代码转换的条件和原则，就不难对于汉字终端技术有一个初步概括的了解。

1·4 汉字终端技术的历史和现状

汉字终端技术是一门新兴的多学科综合性技术，它在五十年代末期逐步发展起来，并在近十多年中获得了飞速的发展。

早在1958年，我国电信科技人员为了使我国能用机器直接实现中文通信，曾设计研制过中文电传机的样机。样机的主要部件是一个长约300mm，直径约250mm的铝质滚筒，滚筒里集装着数千个刻有汉字的字头，滚筒能自转和轴向移动。通过五单位凿孔纸条上的不同孔位，由五根选择探针和连杆、再通过凸轮传动到滚筒上所需字头的位置。定位后，由字锤迅速推出字头打印在色带和纸页上。整个样机全靠机械部件传动，外形象一台四方桌，体积很大。限于当时的技术条件，该样机未达到实用水平，这是我国第一台汉字终端的雏形。

1969年，我国研制成功第一台达到实用水平的汉字终端设备——中文电报译码机。机器的核心部件是一个能存储几千个汉字的磁芯记忆装置——字形存储器。该机用五单位凿孔纸条通过光电阅读器输入一个汉字的电码信号后，再通过译码电路转换成该汉字的存储地址，字形存储器输出相应的汉字字形信号到静电印字机，就能印出这个汉字的字形来。中文电报译码机的基本原理与现在的汉字终端较相似。由于使用了晶体管、磁芯等分立元件，因而整个机器的体积很大，电路复杂，成本较

高，只适合在报务量很大的少数中心电报局使用。

随着计算机技术和集成电路技术的发展，经国务院有关部委的倡导和组织，于1974年8月把汉字信息处理技术的研究开发工作正式列入国家计划，定名为748工程。该工程包括精密汉字编辑照排系统、汉字情报检索系统、汉字通信系统和汉字终端设备的研制等内容。具体的研究课题有：汉字输入输出方法及设备、汉字文字处理的基础理论和标准、引进微型计算机系统的改造和专用系统的研制、中西文兼容的系统软件及应用软件等等。1981年6月，成立了全国性的学术团体“中国中文信息研究会”（86年10月改名为“中国中文信息学会”）。此后，汉字终端和汉字信息处理技术在政府部门的组织安排和学术团体的支持配合下，得到十分迅速的发展，取得了一系列的成果。

在汉字输入编码技术的研究方面，十年来方兴未艾。到1987年止，全国（包括台湾省）已有各类汉字键盘输入编码方案约700种。美国和澳大利亚的华裔学者也提出了数十种。86年5月，国家有关部门组织了第一次全国规模的汉字输入编码方案评测，参加评测的输入编码方案34种，由于种种原因没有参加评测、但已经上机实际运行的输入编码方案不少于35种。各种新的编码方案还在不断地产生。一些比较优秀的汉字输入编码方案，经过几年来的实践检验和不断改进完善，正在被迅速推广应用。

近几年来，为了突破汉字输入这个“瓶颈”，愈来愈多的人投入了对汉字识别和语音识别技术的研究和开发。进展较快的印刷体汉字识别技术，已经能识别三、四千个常用汉字，识别正确率在98%以上，识别速度平均每秒一、二十个汉字。手写体联机实时识别输入技术和特定口音的单呼口令语音识别输入

技术，也取得很大的进展，并且已有一些实用性的设备投入使用，其应用前景是十分广阔的。

在汉字设备的研制方面，激光汉字照排系统已于85年初投入实际使用，并获得86年日内瓦国际发明展览会金牌发明奖。汉字显示器、汉字24针打印机在国内已能实现批量生产。特别是近几年来全国各行各业开发研制的各种专用和通用的汉字终端设备和汉字信息处理系统多达数百种。在研究开发各种高分辨力和彩色的显示和打印技术的基础上，各种新型的显示器和印字机正在不断地出现，并且逐步进入实用领域。如大面积液晶显示器，电致发光、等离子体发光显示器；热转印印字机，喷墨印字机，激光印字机和液晶印字机等。各类汉字设备正朝着小型化和高性能价格比方向发展，并已基本达到同类西文设备的技术水平和实用程度。如各种汉字（中英文）电子打字机、汉字电传机、汉字显示终端等，早已丢掉了沉重庞大的外壳，几乎与同类西文终端设备一样小巧灵便。在机关事务处理、企业管理、通信、情报检索和办公室自动化等领域得到越来越广泛的应用。在汉字信息处理的基础理论研究和标准化工作方面，也取得了一系列令人瞩目的成果。近几年来颁布实施的有关汉字信息处理的国家标准主要有：《信息处理交换用七位编码字符集》（GB1988-80），《信息交换用汉字编码字符集基本集》（GB2312-80），《文字和符号成形设备用的增补控制功能》（GB5261-85），《信息交换用汉字（15×16、24×24、32×32）点阵字模集及数据集》（GB5199、GB5007、GB6345）；还审查通过了《2K和4K的"汉字整字键盘》、《信息交换用汉字编码字符集第二、四辅助集》等。这些标准的制定和实施，必将进一步促进我国汉字信息处理技术的发展。

1·5 汉字终端技术的基本内容

汉字终端技术是计算机汉字信息处理技术的一部分。它的基本内容是研究汉字的输入、输出技术，以及为实现汉字输入输出所必需的汉字编码技术、存储技术和应用软件技术，汉字通信技术也是汉字终端技术的一项重要内容。为了实现信息的传输和交换，必须要求实现汉字传输代码标准化，并妥善解决与原有的西文字符通信的兼容问题。

1. 汉字输入技术

汉字输入技术要解决的问题，是如何实现从汉字字形到机器内部码的转换。概括地讲，实现这一转换可有两种不同的途径。一种途径是由计算机模拟人的智能自动识别汉字。如用光学字符阅读机（OCR），将汉字字形直接转换成汉字内部码，用图形输入板将手写汉字转换成内部码等；另一种途径是通过操作各种键盘来输入汉字，也就是由人识别汉字字形并将它转换成输入码，再由计算机将输入码转换成汉字内部码，这是目前将汉字输入机器的主要途径。用西文键盘输入汉字，可不另增加设备投资，因而得到广泛的应用。以西文字符键盘为基础的汉字编码输入技术，已经成为汉字输入技术的重点之一。

2. 汉字存储技术和输出技术

汉字存储技术和输出技术要解决的问题，是将经过机器处理后的汉字代码（内部码）转换成人们能够理解的字形或声音。要把代码转换成字形，就需要有存储字形数据的字形发生器，即汉字库。要把代码转换成声音，也需要有存储语音参数

的存储器。只有把字形或语音参数事先记录在适当的存储介质上，当汉字代码输入时，才能有相应的汉字字形或语音输出，这是汉字存储技术的主要内容。另一方面，需要设法将字形信息在一定的媒体上显现出来，以便于人们阅读或保存，这是汉字输出技术的主要内容；显现的方法有显示和打印两类。打印输出的文本应有一定的格式（如分段分页）和字体区分（如标题字体、正文字体等），汉字文本还有横排、竖排等不同的排列要求。

3. 汉字终端的软件技术

与通用的计算机系统一样，软件是汉字终端技术的重要组成部分。它包括汉字输入和输出管理软件，字库管理软件、汉字编辑软件等内容。汉字终端软件技术中的一个关键问题是如何解决中西文兼容问题。为了使汉字终端既能处理汉字，又能处理字母数字，就要为每个汉字规定一个唯一的机内代码，使它既能依靠编码本身的特点与字符代码相区别，又能在不涉及判别中西文差别的处理过程中表现出与字符代码相一致的特性。

4. 汉字通信技术

汉字通信技术要解决的问题，是如何准确迅速地实现汉字信息的传输和交换。汉字通信就其代码制式、接口、传输和控制规程等方面来讲，属于数据通信的范围。某些汉字电传机则仍保持传统电报通信的特性。汉字通信技术的关键是要规定统一的汉字传输代码及适合汉字特点的控制规程，并妥善解决与原有西文字符通信的兼容问题。汉字通信中的信号传输和交换方式以及纠错编码技术，也是汉字通信技术的重要内容。

二、漫话汉字

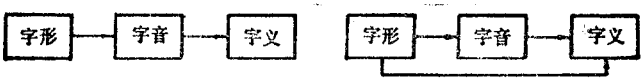
汉字是记录汉语的文字，是我们日常工作、学习和社会交往中时刻不可离开的工具。人们读书、看报、写文章、做记录，都要和汉字打交道。但是，在正常情况下我们并不需要去分析每个汉字本身的特征，而只是把注意力放在对汉字所表达的信息内容的思考上，只是遇到不知其意的生僻字需要查词典时，才去想一想它属于什么部首、有多少笔画，或者怎样读音拼写等。也就是说，当需要使用工具书时，就得简单分析一下这个汉字的字形或字音特征了。现在，当我们需要借助电子计算机这个现代技术工具进行汉字信息处理时，有必要了解一些有关汉字的演变发展和音形特征的基本知识。

2·1 汉字的历史演变及特点

汉字是世界上最古老的文字之一，距今已有六千多年的历史。我们的祖先在劳动和日常交往中创造出来的象形和会意图形，就是原始汉字的雏形。如描绘成象太阳、月亮、流水、高山形状的图画文字，就是“日、月、水、山”等字的原始象形文。我国汉代著名文学家许慎在《说文解字》中写道：“仓颉之初作书，盖依类象形，故谓之文。其后形声相益，即谓之字。文者，象物之本，字者，言孳乳而浸多也。”他的意思是：仓颉造字是先有象形的“文”，以后由形旁和声旁相组合，才从象形的“文”繁衍出数量众多的“字”。

古人称独体为文，合体为字。古代“文”通“纹”，是花纹或图案的意思。独体的“文”依照事物本身的曲线形状来勾画，笔势婉转曲折，结构浑然一体。但是，能够“画成其物”“察而可见”的象形文数量毕竟不可能很多。于是就以这些独体文作为部件，“形声相益”，造出了大量形声字即合体字。在现存可识的最古老的汉字——甲骨文和金文中，已有了象形字和形声字。这种文字的笔画结构并无一定之规，象什么就是什么。秦代统一六国文字，创造了篆书，简化了许多曲线图象，形成了具有某些规律的笔画，但仍大体保留了象形文字的特征。由篆书简化演变而来的隶书，把圆转的笔画改为平直方折的有规律系统，奠定了汉字楷书字形的基础，但字形的基本结构没有变。而且，随着我国特有的书法艺术的发展，汉字也就越来越成为利于书写的方块象形文字了。

文字是记录语言的符号体系，不论是汉字还是拼音文字，都是以一定的字形来记录语音、表达语义的。也就是说，都具有“形、音、义”三个特性，但是，对于不同种类的文字，三者之间的关系也不同。图2·1是拼音文字和汉字这两种文字的“形音义”关系示意图。



(a) 拼音文字　　(b) 汉字

图 2·1　两种文字的“形音义”关系示意图

拼音文字是以字音为主的文字系统，字形是音的符号标记，并不直接联系字义。由于语音的音素数量有限，而且发音总是先后有序的，不可能同时发两个音。因此，人们只需用少量可作为区别发音的记录符号——字母，并以时间先后为序线

性排列，就形成了表达不同含义的文字。汉字则是以字形为主的表意文字，又称象形文字，字形还常常直接联系着字义。一个汉字就是代表一定“形音义”的结构完整的意义单位。在实际应用中，人们除了以不同的字形来区别语音、表达语义外，还能够以字形直接区别和表达语义。

据统计，汉字中除了少数独体字外，绝大多数是合体字。所谓合体字，就是由两部分组合而成的字，其中一部分表示这个字的意义，叫做形旁；另一部分表示这个字的读音，叫做声旁。所以合体字又称为形声字。比如“钢”字，“金”是形旁，表示“钢”是金属一类；“冈”是声旁，表示“钢”读“冈”（gāng）音。由“金”作形旁，和不同的声旁结合，可以组成与金属有关的许多合体字，如“铜、锌、铝”等。由“冈”作声旁，和不同的形旁结合，可以组成“刚、纲”等读音相同的合体字。

了解形声字的构造规律，可以帮助我们认字和记忆，还可以根据形旁表意的特点，从一个字的形旁推断它的含义。比如“瞻仰”的“瞻”字，形旁为“目”，可以推断与眼睛有关；“赡养”的“赡”字，形旁为“贝”，可以推断与钱财有关（古代以贝为货币）。但是，由于汉字使用的年代很久，古今语音变化很大，靠声旁来推断字音，有时往往不准确。比如“地、池、他、施、拖”声旁都是“也”，可是现在一个都不读yě；“济、霁”的声旁是“齐”，可是现在都读jì不读qí。有些形旁今天也已经不能表意了。比如“褒”“衷”，古代都和衣服有关系，所以用“衣”作形旁，在现代汉语里，这些字与衣服没有什么关系，从“衣”旁推不出字义来。据统计，现代汉语用声旁表音功能的只有39%。换句话说，如果按照“长字认一截，短字认半边”的老规矩办，那么就会有61%的字要读错。

我们了解和掌握汉字的历史特点及构造规律，不仅仅是为了学习和应用它，而且要用现代技术手段来对它进行处理。

从学习和应用方面来看，汉字与拼音文字相比，可说是各有所长，也各有所短。拼音文字基本上“字话一律”，具有“见字即能念，会说即能写”的优点。但是，正由于“字话一律”，有多少话就要有多少字，常用的字量是非常大的，这就大大增加了教育识字的难度；汉字能够“以形表意”，音同字不同，便于以字形来辨别同音字，减少了常用字量，还能使不同方言地区的人们用统一的文字交换信息。汉字的单音词特别多，一个单字代表一个语素，几个单字灵活运用，具有极强的构词能力，这些都是汉字的长处。但是，从应用现代技术进行文字信息处理的角度来看，尤其在实现对机器的输入输出方面，汉字要比拼音文字困难得多。

拼音文字的字数再多，总是由26个（或稍多几个）字母拼成的。因此，只需用几十个键的小型键盘，按照文字本身的拼写次序顺序击键，便能直接输入机器；再用只有几十个字模的打印装置，便能实现用机器写字代替手工写字。根据这一原理，西方人在上个世纪就发明了打字机。由于机械结构简单，使用方便，很快得到普及推广，成为使用拼音文字人们不可缺少的书写工具。随后，又发明了电传打字机。在本地打键盘，远方（经过通信线路）便能立刻打印出文字来，使文字信息的传递极为迅速方便。汉字的常用字数虽然比拼音文字要少得多，但每个单字表现为结构完整和相对独立的整体，不易于拆分成类似字母的结构单元。因此，过去仿照西文打字机的原理设计制造的机械式中文打字机，设有几千个汉字的整字字模，只能供专职操作人员使用，始终未能走出打字室的狭小天地。随后试制的机械式中文电传机，也因为体积庞大，价格昂贵而没

有投入实际使用。本世纪中叶以来，电子计算机和微电子技术的发展，把人类推进到信息社会。电子计算机在信息处理领域里的广泛应用，扩大和延伸了人的识别、记忆、判断和执行功能，大大促进了知识资源的开发，使人类文明跨进一个崭新的阶段。西方人用了近四十年时间，发展了一套成熟的技术，成功地实现了适合拼音文字特点的计算机信息处理。毫无疑问，我国的科技工作者应当充分吸取和运用世界先进的科技成果，发展适合我国文字特点的计算机信息处理技术，使古老的方块汉字跨过机械式打字机的阶段，直接进入现代信息社会的大门。

2·2 汉字的数量及使用频度

汉字在几千年的历史长河中历经演变发展，积累的总字数大约有五万六千多个。汉字总量很多，但它们的使用频度是极不均匀的。

据汉代《说文解字》记载，当时的汉字累计有九千五百多个，加上异体重文，总数约有万余个。以后汉字字数逐年增多，发展到清代《康熙大字典》收字四万七千多个；民国初年《中华大字典》则收字四万四千多个。目前正在编纂并已开始陆续出版的《汉语大字典》收集古今汉字共约五万六千余个。在这么多汉字中，绝大多数（四万多个）已经淘汰不用，称为“死字”，只有在整理古籍等特殊场合才可能遇到这些陌生的“面孔”。若以现代编撰的《辞海》和《现代汉语词典》为参考，现代可能用得着的汉字，包括异体字和繁体字在内，约有一万一千到一万四千余个。这里说的“可能用得着”，是相对于那些已完全用不着的“死字”而言。实际上，在这一万多

个汉字中，约有一半左右（约七、八千个），在现代文献中已经罕见。在一般应用范围内通用的汉字总数约为六、七千个。1975年和1985年，我国曾进行过两次规模很大的字频统计，统计范围包括时事政治、社会科学和自然科学的各个领域。统计结果表明，在几千万字文献资料中，实际用到的汉字单字总数为6335个。在这六千多个汉字中，有三千多个字的累计使用频度达到99.9%以上，另外三千多个字的累计使用频度则不到0.1%。人们把前者称为常用字，后者称为非常用字或次常用字。

统计结果还表明，常用字也是“忙闲不匀”的，它们的使用频度相差悬殊。在这三千多个常用字中，二千四百多个字的累计使用频度达99%；一千个字的使用频度已达90%；其中有163个最常用的“高频字”，其累计使用频度高达50%；而“的、一、是、在、了、不、和、有”等八个字的累计使用频度竟达10%，也就是说，平均每10个汉字，就有其中的一个。

综上所述，汉字总量庞大，而使用频度相差悬殊，那末，汉字终端应选多少汉字为宜呢？这需要从经济和实用的角度来考虑。目前国内通用的汉字终端能处理的汉字数，是以国家标准《信息交换用汉字编码字符集，基本集》即GB2312-80所收集的6763个字为准。该标准把6763个字分成两级，第一级收集常用字3755个，统称为国标一级字；第二级收集次常用字3008个，统称为国标二级字。有些简易型的汉字终端设备则只选用国标一级字3755个。

应当指出，这里所谓的常用字与次常用字是以面向各个应用领域的平均使用频度区分的，是指通用范围而言。在不同的专业范围应用时，情况可能会有很大的不同。例如：“髁、髀”，“霰、霾”，“嘀、酊”，“帷、幄”等字，在通用范

围内的使用频度并不高，故列入国标二级字中。但是在医学、气象、生化、军事等专门领域中就可能会用得较多，而国标一级字中的许多字则可能是不常用的。因此，在设计和使用汉字终端设备时，恰当选定能够处理的汉字数量，并对不同使用频度的字加以区分，可以提高工作效率和降低成本，有利于设备的普及应用。

2·3 汉字的字形特性

字形是汉字“形、音、义”特性中最重要的特性。了解汉字的字形特性，对学习和应用各种汉字字形编码输入方案很有必要。

上面提到汉字数量成千上万，主要是以不同的字形来区别的。特别是在汉字终端技术的范围内，字形是区分不同汉字的唯一依据。这就是说，同一字形的，都被认为是一个字，都对应唯一的国标码和内部码，当然也对应唯一的字形码。例如“长”字，不论是“长期”的“长”（cháng），或者“生长”的“长”（zhǎng），都作为一个字，对应国标区位码19-04。又如“银行”的“行”（háng），或者“行军”的“行”（xíng），都作为一个字，对应国标区位码48-08，尽管它们的字音和字义都是不同的。反过来说，对于字音、字义相同但字形不同的繁体字、异体字，就被作为不同的字加以分别处理。

众所周知，汉字的字形结构繁简不一，除了极少数几个笔画最简的字以外，绝大多数汉字的字形是可以分解的，它们的结构层次可用图2.2来表示。

由图可知，汉字的字形结构可概分为单字、部件、笔画三个层次。它们的构成规律是：由笔画构成部件，由部件构成单

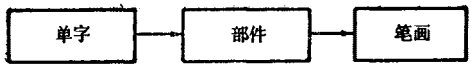


图 2·2 汉字字形结构层次示意图

字。从层次关系上讲，单字是表层结构、部件是中层结构，笔画是底层结构。从数量关系上讲，几十种笔画构成几百个部件，几百个部件构成几千个汉字。建立汉字字形层次结构的概念，有助于深入地认识汉字的字形特性，这是理解各种字形编码法的基础。

现以表2·1来具体分析“中、国、汉、字”四个字的字形结构特点。

从表2·1可见，单体型（如“中”）字的字形结构紧密，部件拆分的直观性差，人们在查字典时也往往难以确定属于何类部首；上下型（如“字”）和左右型（如“汉”）的合体字，部件拆分比较方便、直观。但是，部件拆到什么程度适当？特别是对一些字形结构比较复杂的字，仍需要加以认真斟酌。

下面对字形结构的笔画、部件、单字（字型）三个层次分别加以讨论。

1. 关于笔画

笔画是写字下笔后按照一定笔向连续画成的一笔，是汉字字形结构的最低层次。笔画具有既有联系又有区别的三个属性：笔画数、笔画形状（笔形）和笔画顺序（笔顺）。

笔画数即笔画的数量，这是人们很熟悉的概念。由于一个楷书汉字的笔画数是确定的，也比较容易掌握，因此字典排序或人名排序常用笔画数作为先后次序的标准。

表2·1　　汉字的字形结构

单字	字型	部件		笔画	
		形状	名称	形状	名称
字	上下型	宀	宝盖头	丶	点
				丶	点
				𠃌	折
		子	子	𠃌	折
				亅	折
				一	横
汉	左右型	氵	三点水	丶	点
				丶	点
				丿	撇
		又	又	𠃌	折
				乚	捺
国	内外型	口	口	丨	竖
				𠃌	折
				一	横
			玉	一	横
				一	横
				丨	竖
				一	横
				丶	点
中	单体型	口	口	丨	竖
				𠃌	折
				一	横
		丨	竖	丨	竖

汉字笔画数的简繁差别是很大的。笔画数最少的“二”和“乙”字，只有一笔；大多数汉字的笔画数在7画到13画之间；而字形结构复杂的字，笔画数在30画以上。据统计，国标汉字基本集的6763个汉字的平均笔画数为10.6画，《辞海》中16339个汉字的平均笔画数为12.7画。因为后者收集了大量繁

体字和异体字的缘故。

在一般情况下，区分一个汉字的笔画数是比较容易的。如“王”是四画，“田”是五画等，人们都很熟悉。但在某些情况下，即使是常用字的笔画也往往会搞错。例如，常见的“草字头”（艹）和“走之旁”（辶），正确的写法是三画，但有些人还习惯于写成四画（艹和辶），这类属于新旧字形的变化而引起的笔画数改变，在词典的“新旧字形对照表”中可以查到。这里着重谈谈正确区分一个汉字笔画数的关键，就是如何处理多折笔画的连笔或断笔问题。

对于在字形结构上首尾相接的多折笔画，区分其笔画数的一般规律是：在楷书书写规范条件下，凡能连续保持从左到右、从上到下的运笔方向，均作为是一画。例如角形笔画结构“乚、㇆”和“亅𠂆”，前者在书写时能保持从左到右和从上到下运笔方向，故为一画；后者中间要断笔一次，否则就会出现从下而上或从右至左运笔，故为二画。以此类推，多折笔画“㇊、㇋、㇅、乙”，在书写能保持连笔，都是一画；写“𠃊”时要断笔两次，故为三画。又如方形笔画结构中的“凵、コ、冂、匚”四种开口式笔画结构，都需中间断笔一次，因此都是二画；对于封闭的笔画结构，习惯的书写顺序是最后一笔封口，所以“口”是三画（即丨㇆一），而不是两画；“凸”是五画（丨一丨㇈一），而不是四画；也就是说，封闭结构的左下角的角形笔画结构不能写成一笔（乚），而是两笔。至于带“提”或带“钩”的笔画，不论其向左或者向右（如“亅”、“𠄌”），均按连笔的一画处理。

笔形是一笔画的形状。笔画数和笔形是汉字笔画的两个相互联系的特性。笔画数表现为一个抽象的数目；笔形则表现为一个具体的形状，其变化繁多复杂。分析汉字字形，主要是分

析笔形及其具体的结构关系。

分析汉字笔形，传统的书法家以“永”字为例，根据书法艺术的特点提出了“永字八法”。今天人们根据硬笔写字的运笔方向归纳简化成“横（一）、竖（丨）、撇（丿）、点（包括捺、㇏）、折（乙）”五种基本笔形，即所谓的“札字五笔”。这种归类法常见于各种字典的部首排序及一些常用的笔画编码法中。有的分析汉字笔形归纳为六类，就是把“折”笔分成“左折”（𠃌，又称“弯”）和“右折”（乚，又称“拐”）两类，就成为“横、竖、撇、捺、弯、拐”六类笔形；还有的在上述六类笔形的基础上，又增加了“叉”（+）和“方”（口）两种复合笔形，就成为“横、竖、撇、捺、弯、拐、叉、方”八类笔形了。这后两种笔形分类方法也常见于一些笔画编码法中。人们把笔画分得细一点，有利于简化输入编码及减少重码字。

在笔形分类中，不论何种方法，通常都是把“挑”（㇀）笔归入“横”笔；“点”（㇏）笔归入“撇”笔。这是因为：第一，它们的运笔方向相同，“横”和“挑”都是从左到右，“撇”和“点”都是从左上到右下；第二，它们在许多情况下常互相转化。例如“土”的末笔为“横”，但作为偏旁时末笔就变成“挑”了。还有“工、王、立”等字都是这种类型。又如“木”字的末笔为“撇”，作为偏旁时末笔就变成“点”了。还有不少同样的字例。所以，这样归并是比较合适的。

现将五种笔形的主要特点及举例说明列于表2·2。

笔形和笔画数一样，也常用于汉字的排序。在查字典和索引时，对相同笔画数的部首或字，常以起笔的笔形为序。起笔的笔形也相同时，以第二笔的笔形为序，依此类推。笔形排序的规律有两种，一种是按“一、丨、丿、㇏、乙”为序，《辞

表2·2　　五种笔形的主要特征及举例

种类	走向	笔形	举例说明	出现频度
一	从左至右	一（横）	一　工　上	30.46%
		㇀（略向上）	七　（第一笔）	
		丿（挑）	工（工字旁）　土（上字旁）	
			扌（提手旁）　氵（三点水）	
丨	从上至下	丨（直）	上　下　中	19.62%
		亅（略向左）	五　（第二笔）	
丿	从右上至左下	丿（斜）	人　大　木	15.99%
		亅（较直）	儿　月　久　用　（均第一笔）	
		丿（短）	匕　千　反　鸟　（均第一笔）	
乀	从左上至右下	乀（斜）	入　木　水　（均末笔）	16.74%
		㇏（较平）	之　辶　廴　（均末笔）	
		丶（点）	宀（宝盖头）　氵（三点水）（第一、二笔）	
乙	弯拐曲折钩	𠃌（左折）	丑　马　刀　也　（均第一笔）	17.68%
		𠃍（右折）	无匹　（末笔）互　（第二笔）	
		𠃋（多折）	乃乙及（第一笔）讠（第二笔）	
		𠄌（带折）	一（平宝盖）予（第一、三笔）	
		亅（带钩）	丁戈才（提手旁）犭（反犬旁）	

海》及多数字典采用这种排序法；另一种是按“丶、一、丨、丿、乙”为序，《现代汉语词典》及有些字典采用这种排序法。现行汉字编码法的笔形排序两种都有，但前者使用比较普遍。

如前所述，汉字的笔形具有鲜明直观的特点，是识别和区分汉字的重要依据。人们要知道一个汉字的笔画数，必须一笔一笔写完，而对一个字的笔形常常可以一看便知，一目了然。

人们在阅读一般的书报杂志时，常有快速浏览一下便可知其大概内容的体会，有的还能达到所谓“一目十行”的地步，这大概是汉字笔形变化鲜明直观带来的一大特色吧！这同时也产生了一个副作用，就是人们在着重了解字义的时候，对字形变化的一些细微末节不大注意。在用手写字时往往就一笔带过，不去深究了。可是，要用笔形编码法向计算机输入汉字时，就需要有点认真精神了。稍有一点马虎，计算机就会“不予理睬”，甚至闹出“张冠李戴”的笑话来。这里举一些易于搞错的例子。如“丑”字的中间（第三笔）是一个“横”笔，不是“点”笔；“草字头”（艹）在一横笔上挂的是两个短“竖”，不是一“点”和一“撇”；“凯旋”的凯字左下角是“己”形（三笔），“一卷书”的“卷”字和“模范”的“范”字的右下角是“已”形（两笔）；“吴”字“口”下面的第一笔是“一”不是“折”（㇕）等等。这些在国家有关部门正式颁布实行的《简化字总表》中以及一些字典卷首的说明中都有注明，我们在学习和应用时稍加留意，便可以避免出错。

最后谈谈笔顺。笔顺是人们写字时的笔画先后顺序。按照一般的书写习惯，大多数字的笔顺已经约定俗成，没有疑义。总结笔画顺序的规律，大致有以下十条：先上后下，如“合”；先左后右，如“心”；先中后旁，如“小”；先外后内再封口，如“同”、“田”；先整后零，如“戈”；先横后竖，如“干”；先横后撇，如“夫”；先撇后捺，如“义”；先撇后点，如“丸”；先钩后挑，如“扎”。

应用汉字的笔顺属性，可以很方便地把具有二维图形结构的方块汉字变换成若干基本笔画的线性排列，就象用若干字母排序成为拼音字那样。根据这一设想设计的几种笔画笔顺编码法，由于符合大多数人的书写习惯，用类似打拼音文字的方法

输入汉字字形，规则比较简单，极易学习掌握。这里的困难问题首先在于有些字的笔画顺序还不统一，尤其许多常用部首和单字笔顺不一致的情况还比较多。上面提及的十条一般规律，也很难在书写时同时遵守不悖。例如，常用的“竖心旁”（忄），习惯写法就有三种笔顺：点、竖、点（先左后右）；竖、点、点（先中后旁）；点点相连再一竖，便于连笔书写。又如“兆”字，有的按照“先中后旁”的规则，先写中间的“撇”和“钩”；也有的按照先左后右的规则，先写左边的“点”和“挑”。在实践中这样的例子是不少的。还有一个困难，笔顺是属于写字过程中的一个动态参数，在已经写成的字面上是看不出来的，也不能从词典之类的工具书中查到明确的答案。因此，在遇到疑义时，只有参考讲解笔画笔顺的专门书籍方能解答，况且，目前尚无公认统一的笔顺标准。

2. 关于部件

部件又称为字根、字元、构件等等，是汉字字形的基本构成单元，是字形结构的核心。在汉字字形结构的三个层次中，部件作为一个中间层次，具有承上启下的关键作用。从汉字形成过程的历史演变和造字规律来看，部件也是汉字字形结构的基础。几百个部件构成了今天成千上万个汉字。有一副对联：”此木为柴山山出，因火成烟夕夕多”，就是部件构字特点的一个生动形象的写照。

那末，部件构字有哪些特点和规律呢？据武汉大学的汉字结构分析统计系统（CCIAS）的统计资料，在常用字范围内，绝大多数汉字是由2～4个部件组成，其中以3个部件组成的汉字数量最多。单体字（单部件结构）和5个部件以上的复杂合体字数量只占极少数。构成《辞海》全部汉字的部件总

数为675个；构成国标基本集汉字的部件总数为564个；构成《新华字典》全部汉字的部件总数为500个。平均每个汉字所用的部件数为3个左右，详见表2·3。

表2·3　　汉字部件总数及平均部件数

汉字集名		《辞海》	国标基本集	《新华字典》
部件总数		675	564	500
每字部件数		3.737	3.205	2.94
单字	1	2.19	4.76	3.6
构字	2	16.54	23.72	25.6
部件	3	30.18	36.74	46.4
数分	4	24.43	21.28	22.5
布(%)	≥5	26.65	13.49	1.9

据以上分析，构成全部汉字的部件总数有五、六百个。但它们的组字频度差别很大。有200多个部件，就已经占构成汉字总数的80%以上。其中，部件“口”的组字频度最高，约为20%左右，就是大约有五分之一的字要用到它。在使用中，还有一个“实用频度”的概念。实用频度是部件在全部使用的汉字中出现的频度，是不同使用频度的汉字对部件组字频度的加权。例如，“白”、“勺”两个部件的组字频度并不高，但由它们组成的“的”字的使用频度高达4%，所以这两个部件的实用频度也上升到前10名之列。现将《辞海》字集的部件组字频度和实用频度的前10个部件顺序列出，见表2·4。

分析和应用汉字部件的构字频度，可以提供部件优选和归并的理论依据，做到用尽可能少的部件组成尽可能多的汉字。

最后讨论一下部件的笔画结构方式及类别。如前所述，部件是由笔画构成的，从形体关系分，根据笔画结构的不同方

表 2·4　　常用部件的组字频度和实用频度

序　号	1	2	3	4	5	6	7	8	9	10
组字频度	口	一	人	艹	日	木	亠	亻	氵	土
实用频度	口	一	人	土	日	亻	白	勺	八	又

式，可以分为单、散、连、交、混五种部件结构类型。所谓“单”就是单笔部件。上面提到的五种基本笔画，有时也充作部件，就称为单笔部件。其它各类部件的类型及结构方式举例见表2·5。

从音义关系分，部件可分为成字部件和不成字部件两种。所谓成字部件，是指这个部件本身就是一个汉字，如“日、月、水、火、人、口、土”等等；不成字部件本身不能单独成字，只作为一个组字的部件（部首、偏旁），如单人旁（亻）、反犬旁（犭）、友字头（广）、三点水（氵）等等。许多不成字部件是从古代的独体文演变而来的。

表2·5　　部件类型及笔画结构方式

部件类型	笔画结构方式	举　例
单	单笔结构	一丨丿乀乙
散	离散结构	三八犭氵心灬
连	相互连接结构	工厂人刀弓尸白
交	相互交叉结构	十×艹丰力车
混	有连有散	幺矛贝讠竹
（合）	有交有散	寸义戈米
（合）	有连有交	土大木王禾
（合）	有连有散有交	舟鱼酉雨

应当说明，由于汉字字形结构的复杂性，加上笔形变化的多种多样，部件的划分和确定目前还没有一个公认的统一标准。尤其在汉字编码的应用范围内，出于对编码效率等多种因素的考虑，部件划分的灵活性比较大。一些笔画结构在较复杂的字形中，可以拆成为一个部件；但在单独成字时，则还需将它拆成为单笔部件和较简单的其他部件。例如常用的“的”字，可以拆分为“白”、“勺”两个部件；但“白”作为单字时，则还要拆分为“丿”和“曰”两个部件。成字部件“大”，在许多字形结构中作为一个部件，但单独成字时则需分解为“一”和“人”两个部件。因此，在初步了解掌握了部件构字的一般规律以后，还需根据不同编码法的规定进行具体分析，方能达到熟练运用的目的。

3. 关于字型

字型是汉字形体结构的基本拓扑图形，是字形结构层次中的表层结构——单字的结构类型。建立字型的概念，有助于人们更深一步地认识汉字的字形特性，从而正确地把单字分解为部件。

如前所述，单字是由部件构成的。以构成单字的部件排列位置来区分，汉字大体可分成四种类型，即单体型、左右型、上下型和内外型。后三类又可合并称为合体型。由于构成单字的部件数量不同，每类字型分别有单根、二根、三根、四根等不同结构，归纳如表2·6所示。

据统计，在汉字字型的四种类型中，左右型结构是大多数，约占汉字总数的63%；上下型结构次之，约占汉字总数的22%；而单体型结构最少，只占2%；余下的为内外型结构，占13%。为了便于阅读，人们给字型中不同位置的部件赋以不

表2·6　　字形结构表

字型		图字	单根	二根	三根	四根
单体型		图式	□	⊠	⊠	
		字例	三月	天电必	直事乘	
合体型	左右型	图式		口	口口口	口口口
		字例		的仆江	结到湘	湖疑溜
	上下型	图式		日	口口口	口口口
		字例		字另召	盟花要	樊照葛
	内外型(及其它)	图式		口口口	口口口	口(多根)
		字例		囚旭区	司这网	履氲溺

同的名称。如："左旁右边"，"上头下底"，"外框内心"，"中腰四角"等等。比如"好"字，是女字旁、子字边；"花"字是草字头、化字底；"因"字是国字框、大字心；"衷"字是文字头、中字腰、衣字底。

怎样才能把单字正确分解为部件？合体字的分解比较容易掌握，一般采用层次二分法（又称树形分解法）进行分解。层次二分法的示意图如图2·3所示。分解的原则可归纳为一句话："从形分解、兼顾音义"。因为汉字是形音义的统一体，兼顾音义符合字形结构规律，也便于理解和记忆。例如"湖"字，第一步应该分解为"氵"（形旁）和"胡"（声旁），而不应分解为"沽"和"月"；"鸿"字第一步应分解为"江"和"鸟"，而不是"氵"、"工"、"鸟"等。

单体型汉字的字形结构紧密，分解比较麻烦。通常可以把单体字看作是已知部件的嵌套结构。分解时，应按照书写顺序

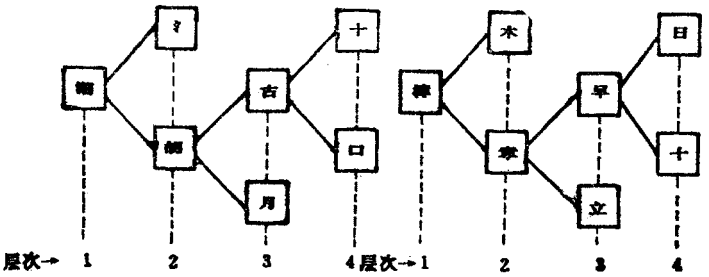


图 2·3 合体字层次二分法示意图

首先抽取以首笔形成的最大部件，然后依次类推。所谓“最大部件”，就是再增加一笔就不成为部件了。当然，也应该考虑到下一个部件的抽取方便。现举例说明见表2·7。

表2·7 抽取首笔最大部件

单体字	首笔部件	其余部件
不	一	卜
生	丿	主
重	千	日 二
来	丰	丶丿 八
果	日	木
妻	一	彐 女
自	丿	目
事	一	口 彐 丨

以上介绍了汉字字形分解的一般规律，在运用时还需根据不同编码法的具体规定（如规定了多少个部件等），才能实现

正确的分解。

2·4 汉字的字音特性

文字是记录语言的符号系统。汉字的“字音”来源于汉语的语音，是汉字的又一重要特性。如前所述，汉字的字形构造复杂多变。相比之下，汉字的字音则比较简单，而且很有规律。因此，在汉字终端中应用字音特性的编码方法，一般比较容易掌握。

众所周知，汉字的字音是单音节的，一个字一个音节，具有“声、韵、调”三种属性。大多数汉字的读音含有一个声母和一个韵母，只有极少数字音只有韵母而没有声母。汉字的音节共有四百多个，加上声调为一千三百多个。要用这些音节来表示几千汉字的读音，一音多字（即同音字）的现象非常突出，如何将它们区分开来，这是各种字音编码方法面临的主要问题。

下面分别简要介绍一下汉字字音特性的声、韵、调特性，以及同音字和多音字。

1. 声母

声母是大多数汉字字音开头的音。例如“他、发”两个字音，“他”的开头是“t”（特）音，“发”的开头是“f”（佛）音。这“t”、“f”就是声母。汉字字音的声母有23个，如表2·8所示。

2. 韵母

韵母是大多数汉字字音除去声母以后的声音。仍以“他、

表2·8　　声　　母　　表

声母及注音	发音特点
b（玻）、p（坡）、m（摸）、f（佛）	嘴唇用力
d（得）、t（特）、n（讷）、l（勒）	舌尖用力
g（哥）、k（科）、h（喝）、	舌根用力
j（基）、q（欺）、x（希）	舌尖下垂
zh（知）、ch（蚩）、sh（诗）、r（日）	翘起舌尖
z（资）、c（雌）、s（恩）	舌头平伸
y（一）、w（屋）	短　音

发”两字为例，除去声母“t”音和“f”音以后，剩下的“a”（啊）音，就是韵母。汉字字音的韵母共有34个，根据其不同的发音特点及在拼音中的位置，可分为单韵母、复韵母、鼻韵母和介音四类。

单韵母是一个单纯而响亮的声音，发音时必须颤动声带。单韵母共有6个，它们是：

a(啊)、 o(喔)、 e(鹅)、 i(衣)、 u(乌)、 ü(迂)

复韵母是由两个单韵母复合而成的，共有8个，它们是：

ai(哀)、	ei(欸)、	ao(熬)、	ou(欧)
ie(椰)、	üe(约)、	iu(优)、	ui(威)

鼻韵母是在单韵母后面带一个鼻音做尾巴，共有9个。按照鼻音的发音部位不同，鼻韵母分为前鼻音和后鼻音两种，下面所列第一行是前鼻音，第二行是后鼻音。

an(安)、 en(恩)、 in(音)、 un(温)、 ün(晕)
ang(昂)、eng(亨的韵母)、ing(英)、ong(轰的韵母)

介音介于声母和韵母之间，从语音理论上讲，介音是韵母的一部分。在拼读有介音的音节时，应采用声、介、韵“三拼连读法”。介音字母由 i 、 u 、ü充当，带有介音的复韵母共有11个，它们是：

ia(呀)、 iao(腰)、 ian(烟)、 iang(央)、 iong(雍)
ua(蛙)、 uo(窝)、 uai(歪)、 uan(弯)、 uang(汪)
üan(冤)

3.声调

声调是字音高、低、升、降的“调子”。汉语是有声调的语言，语音有了声调可以起到区别语义的作用。例如“踢”和“提”的声韵母都相同，由于声调不同，意思就不一样了。

汉语语音有四个声调，就是阴平声、阳平声、上声、去声，合起来叫做“四声”，它们的符号是：－（阴平）、ˊ（阳平）、∨（上声）、ˋ（去声）。四声之外，还有一个“轻声”，没有符号表示。声调的发音特点及标调举例如下：

字音及标调符		声调名	发音特点
妈	mā	阴平	又高又平
麻	má	阳平	先低后高
马	mǎ	上声	先降低后升高
骂	mà	去声	先高后低
吗	ma	轻声	又轻又短

4. 关于同音字和多音字

如上所述，现代汉语有四百多个可读音节，加上声调后为

一千三百多个音节，称为音调节。用这些音调节作为几千个汉字的字音，一音多字的同音现象就比较普遍。据统计，在《现代汉语词典》中，音节“yi”共有164个汉字。分开声调后，去声“yì”就有89个汉字。同音字数量在一百个左右的还有“jī”、“shi”、“yu”、“fu”等音节。在国标汉字基本集（GB2312-80）6763个汉字中，没有同音字的只有18个。加上声调，也只有262个字没有同音字。同音字的比例高达95%以上。有一篇著名的同音字短文，叫做《石室施氏食狮史》，全文一百多个汉字都是一个“shi”音节，可说是同音字的典型例子。

如前所述，在汉字终端中是以不同的字形来区别汉字的。因此在采用各种拼音编码法向机器输入汉字时，需采取必要的技术措施，才能达到正确输入的目的。否则，机器将无法正确处理。

汉字中除了大量同音字外，还有不少多音字，这类字有不止一个读音，不止一个意义，故又称多音多义字。比如“盛”字有两个读音，“茂盛”的“盛”读shèng，是“多”的意思，“盛东西”的“盛”读chéng，是“装入”的意思；“差”字有好几个读音，“差别”的“差”读chā，“出差”的“差”读chāi，“成绩差”的“差”读chà，“参差不齐”的“差”读cī，它们的意义各不相同。据统计，汉字有七百二十多个多音字，其中一字两音的有六百四十多个，三音以上的七十多个。

多音字在机内处理时，表现为编码的二义性，就是有两个或更多的输入代码对应于同一个汉字内部码和字形码（点阵数据），但对汉字输入操作没有影响。相反，有些编码方案为了方便输入汉字，还特意安排一个汉字有多个输入代码，称之为容错码。就是当操作者对某字的输入代码搞错时也容许输入的意思。

三、汉字输入技术

3·1 汉字输入技术综述

汉字输入是人们把需要处理的汉字信息输入到计算机中去，这是汉字终端和汉字信息处理的关键技术之一。

众所周知，人们所熟悉的文字、图象、声音等信息是模拟信息，计算机内部处理的则是数字信息。把模拟信息输入计算机的方法有很多种，基本上可以分两大类：一类叫做自然识别，或称为模式识别。就是由机器自动识别模拟信息，并把它变成数字信息输入；另一类是键盘输入，就是由人将文字信息变成数据（键盘的座标值）输入机器。显然，键盘输入不适用于声音和较复杂图象的输入。

自然识别是一种方便、理想的先进输入方法，它应用计算机的智能特点，对自然的模拟信息进行鉴别和分类，从而将任何复杂的图形、图象、文字或声音输入机器进行处理。用于输入汉字信息的自然识别装置称为光学字符阅读机或汉字阅读机。它应用光电扫描设备“阅读”汉字，通过信号接收和前处理过程变成数字信息。然后经特征抽取再与识别字典（样本）相比较，从而得出识别结果。根据不同的识别对象和识别方法来区分，汉字自动识别输入可分为印刷体汉字识别、手写体汉字识别和手写体汉字实时（联机）识别三种。应用印刷体和手写体汉字（非实时）识别，能把已有的大量文献、资料、手稿自动高速地输入计算机，是建立大容量的文件数据库不可缺少的重

要手段。如果用键盘输入就不知要耗费多少“人年”的劳动量！应用手写体汉字实时识别输入汉字信息，虽然速度不快，但可免除人们记忆编码规则和操作键盘的麻烦，对于数量不多的汉字输入具有很大的实用价值。

但是，实现自然识别输入的技术难度较大，设备的费用比较昂贵，而且有些关键技术问题尚未得到完全解决。以研究起步较早且最接近于实用程度的印刷体汉字识别来讲，对输入文稿的纸张质量和文字印刷质量的要求比较高，而且一次识别时只能限于一种或几种规定的字体。手写体汉字识别不论是联机（实时）的或非联机的，都对写字规则有一定的限制和要求，目前还不能做到对草写或连笔字体的准确识别。基于以上原因，目前阶段的汉字终端输入汉字信息，仍以键盘输入方法为主。

键盘输入就是人们通过对简单的机械装置——键盘的操作，将汉字信息变成数据输入机器。对于拼音文字，键盘是输入文字信息最廉价、最方便的工具，应用十分普及。由于汉字字形结构繁简不一，不易拆分和排序；加之汉字单字的数量很大。因此用键盘输入汉字信息与输入拼音文字相比有很大的差别。但是与各种自然（模式）识别的设备相比，键盘仍不失为成本便宜、使用灵活方便的一种汉字输入设备，因而得到了越来越广泛的应用。

现用的键盘输入汉字信息的方法有多种，大致可分为整字键盘输入、专用键盘输入和标准键盘输入三大类。

用整字键盘输入汉字，是最早应用的一种输入方法。输入时的操作方法与机械式的中文打字机相似，见字选字、按键输入，具有直观易学的特点。整字键盘一般设三千个左右的常用汉字以及一、二百个字符，还设有若干操作功能键。由于键数

多，盘面面积较大，故通常称为“大键盘”。初期的整字键盘采用键钮开关式结构，一字一键，按键操作比较费力，现已不多见。目前应用较广的整字键盘采用触笔传感方式输入汉字，称为笔触式键盘，又叫做笔触式字盘。因为取消了数量众多的键钮开关结构，使整字键盘的体积重量大为减小。操作时只需用一支特殊的笔或手指触到所需输入的汉字上，机器便能把该字的代码输入。

按照不同的传感方式区分，笔触式键盘有静电耦合式、电磁耦合式、压敏式等多种。笔触式键盘的结构紧凑，操作省力，即使非专业操作人员也比较易于掌握。此外，整字键盘还有换位式（又称主键——辅键式）和翻页式等数种输入方式。

用专用键盘输入汉字，是一种以相对独立的汉字部件（字根）组合拼字的输入方法。它类似于标准键盘输入，但由于设置的键位数较多，拆分和拼合的规则比较简单直观。采用拼字方法还可以解决整字键盘盘面容量有限的矛盾，扩大输入汉字数量。专用键盘的键位数介于整字键盘和标准键盘之间，故又称“中键盘”。

用标准键盘输入汉字，是目前应用最广的一种键盘输入方法。标准键盘又称为字母数字键盘或字符式键盘，也有的称为“QWERTY”键盘、ASCII键盘等。标准键盘的键位只有几十个，除了少数特殊符号键和功能键以外，每个键位的字符排列位置都已标准化。由于目前国内外大量生产的标准键盘已经广泛配置于各种型号的微型计算机、键盘显示器、电传打字机等设备，用标准键盘输入汉字，可以不必另行添置汉字输入设备，降低设备费用，有利于推广使用。由于标准键盘的键位少，用它来输入汉字就必需进行编码。也就是说，要用两个以上的字符（字母、数字、符号）编成代码表示一个汉字。这

就需要进行学习和反复应用才能熟悉，这是标准键盘输入汉字的难点所在。但是，随着各种汉字编码方案的完善、优化和逐步统一，用标准键盘输入汉字的方法必将进一步普及。

现将各种汉字输入方法归纳如图3·1所示.

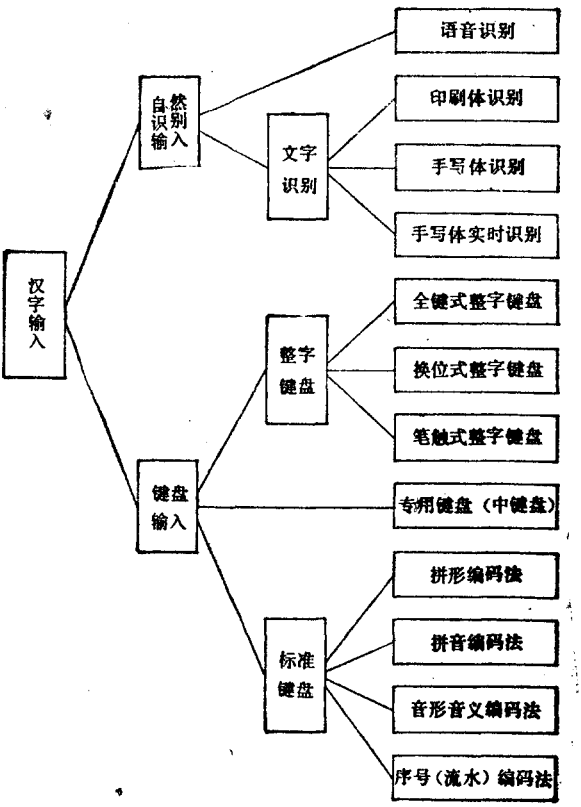


图 3·1 汉字输入方法

3·2 汉字输入编码

1. 汉字输入编码概述

汉字输入编码就是在一定数量的汉字集内给每个汉字编成相应的一组字符代码，以便于用标准键盘输入汉字。

汉字输入编码是汉字信息处理和汉字终端技术的重要内容，它属于信息科学和语言文字学、人机工程学等多学科领域的边缘学科。用于信息处理和通信的近代汉字编码的历史，可以上溯到1881年，我国开通天津到上海的第一条电报通信线路时。当时，为了传送汉字电报，就将汉字先按部首和笔画排序，然后用四位十进制数字依次给每个汉字编号，成为汉字电报码。简称“电报码”或“邮电码”。由于电报码在编码设计时没有在代码结构上反映汉字的形、音、义特性及其内在联系的规律性，而是采用“排队发号”的办法给汉字编码；因此，什么字对应什么码，全凭生记硬背。但是，电报码的编码方法有一个显著的特点，即对于需要编码的汉字集，在每一级分类中对码元符号都是均匀分布的。也就是说，在表示一个汉字的四位数码中，每个码位都可以用 0 ～ 9 十个码元符号来表示。因此它的编码效率高，可以用四位数的等长码组编出一万个汉字及字符的代码，而不会出现一码多字或一字多码的现象，这是其它许多输入编码方案所不及的。

这里我们首先讨论一下汉字编码中的汉字集大小、码元数多少及码组长度之间关系的一般规律。

假设需要进行编码的汉字集有W个汉字，用 n 个码位的等长码组编码，每个码位都可用 m 个码元符号来表示，则它们三者之间的关系可用下式表示：

$$W = m^n \qquad (1)$$

或 $$n = \log_m W = \frac{\lg W}{\lg m} \qquad (2)$$

式中 $m > 1$，是整数。

从上式知，对有W个汉字的字集来讲，所用的码元符号m愈多，则每个汉字码组的码位数n就愈小，码组长度就愈短。或者，对于确定的码组长度（码位数n一定），则所用的码元符号m愈多，可编码的汉字集字数W也愈多。例如在电报码中，码元符号是0～9十个阿拉伯数字（$m = 10$），码组长度是四位等长码（$n = 4$），故能给$10^4 = 10000$个汉字和符号编码。若用a～z 26个字母作为码元符号，用三位等长码组编码，则能表示$26^3 = 17576$个汉字和符号。作为特例，当$m = W$时，码元符号数等于汉字集的字数，则$n = 1$，就是一字一键的整字键盘了。如果$m = 1$，只有一个码元符号，公式（1）、（2）就无意义，编码无法实现。

应当指出，公式（1）、（2）表达的编码汉字集、码元符号数及码组长度三者之间的定量关系，适用于在编码过程中的每一级分类对码元符号都是均匀分布的情况。但上述分析中得到的一般结论，对于其它各种类型的汉字编码方案设计也是有参考意义的。

为了便于输入汉字，在汉字终端和其它汉字设备的汉字输入编码方案设计中，人们总是力求使代码结构与汉字的形、音、义特性建立尽可能密切的映象关系。例如，用字母“O”做码元，来表示封闭形（“口”）的笔画结构；用字母“X”做码元，来表示交叉形（“+”）的笔画结构等等，可以大大有助于对编码方案的记忆和应用。但是，人们在编码实践中发现，不论作为码元符号的字形特性或者字音（字义）特性，在

汉字分类（或分解）过程中出现概率的分布是不均等的，用26个字母难以对应表达错综复杂的汉字笔形，而且会在某些方面具有同一特征的汉字之间出现一码多字（重码）现象。例如，许多汉字的字音特性相同，如果单纯用表示字音的码元符号来编码，就会无法区分不同的汉字，造成重码。为了区分重码字，在码元符号不增加的条件下，就必须在重码字的已有码组上再加一位或几位表示不同特征的码元。无疑，这就加长了码组的长度，降低了编码效率，于是就有不等长编码的出现。所谓不等长编码，就是码组长度n值不是一个固定值。对于那些本来就没有重码的汉字或者使用频度最高的汉字，可以用较短的码组来表示；对于出现重码的汉字码组，则可再增加一、二个附加码元来加以区别。用这种办法可以力求兼顾减少重码和提高编码效率两方面，但是也给输入输出带来一些不便。

在不等长编码的情况下，码长是以码组长度（即每组码位数）的平均数计算的，称为平均码长。按照不同的统计计算方法，平均码长有静态平均码长和动态平均码长之分。

设编码汉字集有汉字W个，每个汉字的码长为n_i（$1 \leqslant i \leqslant W$），则静态平均码长$\bar{L}_s$为：

$$\bar{L}_s = \sum_{i=1}^{W} n_i / W \qquad (3)$$

在实用条件下，每个汉字的使用频度是不同的，设第i个汉字的使用频度为β_i，则动态平均码长$\bar{L}_D$为：

$$\bar{L}_D = \sum_{i=1}^{W} n_i \beta_i \qquad (4)$$

且有：

$$\sum_{i=1}^{W}\beta_i = 1 \qquad (5)$$

因此，在设计不等长编码方案时，总是尽可能把较短的码组赋予常用字；对使用频度低的非常用字则可以安排较长的码组。有些编码方案中，对最常用的高频字专门规定了一字一键的简码或一字二键的缩编码，力求进一步缩短动态码长$\bar{L}_D$。但是，这又增加了人脑的记忆量，降低了编码方案的易学易记性。

再讨论一下多码和重码问题。

信息编码的基本要求是一一对应，但在汉字输入编码设计时，为了兼顾方便使用和提高效率两个方面，在采取了使用不等长码的措施后，一字多码（二义性）和一码多字（重码）的现象仍难以完全避免。如前所述，在汉字输入时，一字多码现象还可以允许。在有的汉字输入方案中，对一些容易写错或读错的汉字，还特意安排了“容错码”，即允许一字多码。但对于一码多字（重码）则不允许，因为这会引起输入信息的不确定性，甚至会造成机器工作的混乱，所以要在编码中采取措施加以区分。

编码字符集中重码字多少的百分比称为重码率。重码率高低是衡量汉字输入编码方案的一个重要指标。重码率有静态重码率和动态重码率两种。静态重码率P_g是汉字集总字数W与重码数C之比的百分数，如下式：

$$P_g = \frac{C}{W} \times 100\% \qquad (6)$$

式中，重码数C＝重码字数－重码组数。

动态重码率是在实际使用中重码字出现的概率，以P_D表示，其定义用下式表示：

$$P_D = \sum_{i=1}^{n} \sum_{j=1}^{m} P_{ij}$$

式中，$i = 1, 2 \cdots n$；$j = 1, 2, \cdots m$；P_{ij}为各重码字组中第 i 组第 j 个字的频度；j 按使用频度高低排序。

显然，动态重码率与汉字的使用频度密切相关。因此在进行输入编码设计时，应把重码字尽可能安排在非常用字的范围内，尽量减少常用字的重码现象，这样可以有效地降低动态重码率。

最后谈谈汉字输入编码的学习和应用问题。学习和掌握汉字输入编码的难易程度，对汉字终端和汉字信息处理系统的普及应用关系很大，是广大用户普遍关心的问题。

应该说明，任何一种有用的知识和技能，都是需要经过学习和训练才能掌握运用的。就以使用标准键盘打字来讲，要用双手十指对几十个键轮流轻快地击键，就需了解字键位置的排列布局，手指怎样分区等知识，并经过一定时间练习才能熟练掌握。如果只满足于用一、两个手指边找边打键的操作水平，那末，即使是一架普通的英文打字机也是难以发挥其应有效率的。今天我们要使用电子计算机进行汉字信息处理或汉字通信，为了把汉字方便快速地输入机器，就得进行学习。因为在使用标准键盘输入的条件下，即使最好的编码方案，也需要有一定的学习记忆量和熟悉操作的过程。当然，面临现在众多的汉字输入编码方案的情况，应该根据各人的年龄、职业等具体条件进行恰当的选择。

从学习掌握的角度看，现有的编码方案可分为三种基本类型：第一种是面向广大使用者的普及型方案。这类方案的特点是规则比较简单易懂，只需较短时间便能掌握操作，但要继

续提高输入速度比较困难。第二种是面向专业操作人员的快速型方案。这类方案的特点是编码规则比较严谨，学习记忆的内容较多，需要经过一定时间的练习方能掌握运用，熟练应用时能够达到很高的汉字输入速度。第三种是介乎两者之间的一般型方案。这类方案吸收前两种方案的长处，兼顾到易学好记和提高操作速度，具有较广泛的应用范围。这三种类型汉字输入方法的学习时间和输入操作速度之间的关系可用图3·2的曲线来表示。图中出现的拐点a、b、c是学习期和使用期分界点。t_1～t_3值的大小是衡量编码方案掌握难易程度的标志之一。g_1～g_3是经过应用熟练后的最高输入速度，它取决于输入一个汉字的平均击键次数和每次击键动作的最短完成时间，以及操作员从看见汉字到获得编码的反应（思考）时间。

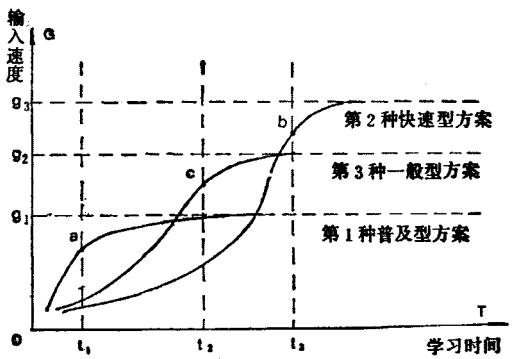


图 3·2　汉字输入编码方法学习时间与输入速度示意图

众所周知，用键盘输入一个汉字的击键次数，取决于该字的码长，即码组的码元数。对于等长码组来讲，输入操作时在字与字之间可以不用间隔键分隔，由机器自动分隔，故击键次数与码长相等；对于不等长码，则必须用间隔键做为字与字之间的分隔标志，因此输入一个汉字的击键次数等于该字的码长加

1。一个汉字输入编码方案的输入汉字平均击键次数，也是衡量该方案优劣的一个重要指标。除了全拼音输入编码和基本笔画编码外，大多数编码方案输入一个汉字的平均击键次数不超过4～5次。根据实验测定，熟练打字员的一次击键动作最短时间约为0.15秒。也就是说，在不假思索地快速盲打的情况下，每分钟可以击键400次。对于码长等于4的等长码方案，最快的汉字输入速度可达每分钟100字。当然，这是在理想情况下推算的结果，在实际操作中，影响汉字输入速度的因素很多。从不利方面讲，汉字编码的输入操作反应时间比较长，即使是最熟练的操作员，也不可能象英文打字或发数码电报那样，可以从“眼见”到“手动”，立即作出反应。通常总需要从汉字到编码之间的思索和反应时间，更不用说在碰到疑难字或非常用字时的情形了；在击键动作时间上分析，长时间操作容易感觉疲劳，就不可能始终保持击键动作的理想速度。从有利方面讲，许多汉字输入编码方案把使用频度最高的几十个汉字编成“简码”，使用2键甚至1键便可以输入，这样能有效地缩短输入汉字的平均动态码长，即减少平均击键次数；从键盘的键位安排上分析，采用键位的合理化布局，尽量使双手的各个手指“各得其所”，轮流均衡地击键。这样可以增强输入操作的节奏感，减轻操作疲劳程度。

综合上述诸因素的分析结论是，如能按照基本指法要求进行操作，完全可以做到平均每分钟输入50～70个汉字，略高于平均每分钟输入40个左右英文单字的速度（据统计，英文的平均击键次数为每字5.2～5.7次）。若在熟练操作的基础上，再大量应用字和词的联想输入，则完全有可能使汉字输入达到每分钟150～300字的高速度。

如前所述，汉字输入编码是涉及到多种学科的综合性边缘

技术学科。在汉字编码、操作人员以及机器的各因素之间，反映在编码方案本身的各个指标之间，存在着互相联系、互相制约的复杂关系。要寻求符合实用要求的最优化的编码设计方案，就需要在现代科学的系统工程一般理论指导下，通过对信息编码理论、汉语语言文字的特点规律、计算机技术以及人机工程学的综合研究，并且在广大群众的大量实践基础上，不断地加以总结提高优化，才能得到真正的解决。

2. 汉字输入编码的种类和特点

如上所述，汉字输入编码已经“出台”的方案有数百种，在一定范围内实际应用的方案也不下几十种。我们可以从编码方案设计时所依据的不同汉字属性，将它们概分为形码、音码、音形（义）码和序号码四种类型。

（1）形码输入方案

所谓“形码”就是把具有一定形状的方块汉字按照某种规则加以拆分、排序、编码。简单地说，就是用拆分字形的方法来编码。形码的汉字输入方案很多，根据它们对字形拆分的不同层次，又可以区分为基本笔画编码和部件编码两类。

基本笔画编码是将字形拆分到预先规定的几种基本笔画，并赋予数字（或字母）代码，按一定顺序排列成码组，来表示一个汉字的输入码。例如“中”字，可以拆分为“竖、折、横、竖”（“丨𠃌一丨”）四个基本笔画，以数字作代码，按书写时的笔顺次序排列，它的输入码就是“2 5 1 2”，或者4个预先规定的其它字符代码。

部件编码是将字形拆分到预先规定的几十种（或一百多种）通用部件，再按照一定的规则将部件转换成有序的代码

组，来表示一个汉字的输入码。例如“忠”字，可以拆分为“口、丨、心”三个通用部件，分别以字母“K、H、N”作代码。如果编码规则只要求以书写时的顺序排列，再没有其它要求的话，则“忠”字的输入码就是“KHN”。由于构成汉字的通用部件数量远大于西文的字符数量，因此用标准字符键盘进行部件编码输入时，编码方案的设计者必须把若干个通用部件归并到一个键位上。也就是需用一个字母（或数字、符号）来代表若干个部件。为了区分由相同部件构成的不同汉字，（例如“口”和“十”可以构成“古、叶、田”等字）通常还需要在输入码中增加一个表示字型的代码。

一般说来，基本笔画编码输入方案的规则比较简单。需用的码元数量（实际上就是基本笔画数量）比较少，只要5～8个数字键的键盘便可以输入。因此，学习掌握比较容易。但是，由于一个汉字的笔画数比较多，基本笔画编码方案的汉字码组比较长，输入汉字时的操作效率比较低。部件编码的输入方案则恰恰相反，由于它的编码规则比较复杂，学习记忆量比较大。但是，部件编码方案的汉字码组比较短，通常不超过4个码元，而且多为等长码组。因此，部件编码方案输入汉字的操作效率比较高，实用性较好，是当前形码输入方案的主流。

形码输入方案的共同特点是直观性，尤其适用于看底稿输入的场合。它不受操作者文化程度高低和各地方言不同的影响，只需看到字形就能按规则击键输入。

（2）音码输入方案

所谓“音码”就是按照汉字的读音，用“拼音”的方法来实现汉字输入编码。音码输入方案类似于拼音文字的输入，是人们比较熟悉的一种汉字输入方法。根据不同的拼音方法，音

码输入方案可分为音素制和双拼制两类。

音素制是直接按照《汉语拼音方案》的规则，用汉语拼音作为汉字的输入编码。也就是说，以26个拉丁字母为码元，每个汉字的码组长度为1～6个字母，是不等长码，字与字之间用间隔键分隔。音素制的输入方案虽然码组较长，输入效率较低，但只要懂得《汉语拼音方案》的人，学会操作键盘就可以输入汉字，不需记背其它任何规则。

双拼制的音码输入方案是把汉字音素加以兼并处理，用类似古代“反切”的方法，不论声母韵母都用一个字母，以声、韵母相拼来表示汉字的字音，故称为双拼法。由于拉丁字母只有26个，声、韵母总数有50多个，所以需用同一字母兼作声母和韵母，只根据它在双拼码组中的位置来区分：位置在前的是声母，在后的是韵母。对只有韵母没有声母的字音，则增设一个“零声母”字符来充当声母。

与音素制相比，双拼制的汉字输入编码方案需要记忆字母与音素之间的替代规则，但是它通常都是等长码，而且码组长度较短，可以得到较高的汉字输入效率。

音码输入方案的特点是适用于无底稿场合的汉字输入。它可以象输入西文那样边思考（或边听）边输入汉字，也不受繁简字体或字形的行书草体变化的影响，在输入常用语句和文章时尤为方便。

但是，由于汉字的同音字多，音码输入方案比形码输入方案的重码现象要严重得多。为了减少重码，人们又设计了拼音词组输入方法。因为汉语中的同音词组要比同音字少得多。不论是音码方案或形码方案，当出现重码字的时候，都需要用“二次选择”的办法来区分。所谓“二次选择”，就是在输入一个码组后，机器将重码汉字以一定的次序全部显示出来，供

操作者选择。操作者只需再按一个数字编号，便可把所需的汉字选中输入机器。

（3）音形（义）码输入方案

音形（义）码输入方案是在音码或形码输入方案的基础上，为了消除汉字输入时的重码字现象而设计的。所谓音形（义）码是指兼顾了汉字的字音、字形（或字义）的汉字输入编码。音形（义）码输入方案也很多，大体可分为以音为主和以形为主两类。

以音为主的音形（义）码，是在音码双拼制选取两个码元符号（即声母和韵母）之后，再增加（后缀）一个或两个表示字形或字义属性的码元符号，借以区分同音的重码字。例如中国的“中”字，用字母“V”代表声母“zh”，字母“P”代表韵母“ong”，它的双拼码是“VP”。为了区别同音的其它“zhong”字，需在双拼码后缀表示字形特征(如起笔或末笔)的符号，便构成音形码。如果后缀表示该字字义特征的符号（常常表现为偏旁部首），便构成音义码。

以形为主的音形码，又称为形音码。它是在拆分字形部件、再选取表示这些部件的码元符号的基础上，再加上这个单字的读音的第一个字母，以区分构字部件相同的重码字。例如“古”、“叶”、“田”三字，都是由“口”和“十”两个部件构成的，在拆分成部件代码以后，再加上它们各自的读音的第一个字母“g”、“y”、“t”，便可以将它们的输入代码分开。当然还有其它区分重码的方法，这里仅是一个举例。

音形（义）码输入方案吸取了形码方案和音码方案的长处，码组是等长码，一般不超过4位，而且重码率很低，避免了输入操作时“二次选择”的麻烦，可以获得较高的输入效

率。但是，有些音形码的编码规则比较复杂，学习记忆量比较大，影响了这类输入编码方案的进一步推广应用。

（4）序号码输入方案

所谓“序号码”就是将待编码的汉字集以一定规则排序以后，依次逐个赋予相应的编号数字（或字母）作为汉字输入代码。用于汉字输入的序号码方案主要有两种，即电报码和国标码。

电报码又称邮电码，是广泛应用于民用通信的汉字传输代码。它以四个数码编成一组，每一组电码代表一个汉字。现用的“电报电码本”是邮电部电信总局1981年修订颁布的《标准电码本（修订本）》，分上、下两篇，各收集汉字约七千字左右。上编按部首排列，下篇按汉语拼音音序排列，上、下两篇的内容相同。

电报码是作为汉字电报通信的传输代码设计的。由于历史悠久，已经约定俗成，因此在汉字电传机及某些汉字终端设备中也用来作为汉字的输入码。

国标码是按照国家标准局1981年制订颁布的《信息交换用汉字编码字符集基本集》（即GB2312）所规定的汉字代码，共收集汉字6763个，分为两级，第一级收常用字3755个，按汉语拼音音序排列，第二级收次常用字3008个，按汉字部首及笔画为序排列。此外还收集了一般符号、序数号、外文字母等图形符号682个。国标码集中的汉字和图形字符，按照它们在代码表中的排列位置，采用两个七位字节编码表示。例如“啊”字，对应代码表中的第一字节是$(30)_{16}$，第二字节是$(21)_{16}$，因此“啊”的国标码是十六进制数“3021”。

在GB2312代码表中的汉字和图形字符，还可以用它们所在的“行”和“列”的位置——即“区号”和“位号”来表示。

例如“啊”字，安排在表中的第16行、第1列的位置上，也就是16区、01位。因此“1601”就是“啊”的国标区位码，简称为区位码。

国标码和区位码是作为汉字信息交换而设计的。码组的代码编成与汉字的音、形、义特征没有多少联系，也是采取“排队发号”的办法编码，因此很难记忆。用它来输入汉字时必需经常查阅码表，比较麻烦。

序号码输入方案的共同特点是它们都有标准的代码表或码本，汉字与码组之间有严格的一一对应关系。用来进行汉字输入时不会产生重码或二义性，因此不需进行二次选择。在熟记常用字的代码以后，可以有很高的输入速度，当然，这通常只有专职的电报报务员或汉字录入操作员才能做到。此外，使用序号码输入时，输入码与机器内部码之间的转换比较简单，甚至可做到两者一致，因此机器处理速度也较快。由于有以上特点，序号码（国标码、区位码、电报码）作为汉字输入方案之一，现已广泛地配置在汉字终端和各种汉字计算机系统上。

3·3 汉字输入编码方案实例介绍

以上我们讨论了四种类型汉字输入编码方案的原理及其一般特点。本节进一步介绍14种应用较广的汉字输入编码方案的编码规则及具体用法，供实际应用时的参考。

1. 基本笔画笔顺编码法

基本笔画笔顺编码法是澳籍华裔学者卢遂现首先提出并设计而成的一种基本笔画编码汉字输入方案。

该方案基于“科学上把一切复杂事物分解到最基本的单元

之后，处理就可以变得简单”的原理，把结构纷繁复杂的汉字字形分解成为“·一丨丿乀”五种基本组成单元。也就是几何学上的“点”和不计长短、只有方向的四种直线线段，它们的方向分别为0°、45°、90°、135°；这五种基本单元类似于“点、横、竖、撇、捺”五种基本笔画，可以用标准键盘的五个字母或数字来表示。例如以数字“1、2、3、4、5”分别表示“·、一、丨、丿、乀”五种基本单元，再以人们写字时的笔画顺序（笔顺）来排序，便可方便地实现任意一个汉字字形的编码。例如汉字“中”，是由方向为0°（水平）和90°（垂直）的5个直线段构成的，其代码就是“32323”。

由此可见，基本笔画笔顺编码法只需用5个数字键（或字母键）便可以输入汉字，而且编码规则十分简单易学。但是，如果只记住以上的简单规则，输入汉字时的按键次数就相当多。例如，常用字“哲”的代码是“231444233232”，因为带钩的竖笔（第二笔）和带折的横笔（倒数第二笔）都是有不同方向的两个线段，要分别取两码。所以输入一个“哲”字要按键12次，显然是很麻烦的。

为了克服这一缺点，提高汉字输入效率，该方案充分利用了标准键盘的几十个键位。除了代表5个基本单元的5个键位外，还设计了31个带不同笔画组合的多画键，使许多常用的部首和笔画组合可以只按键一次便能输入。其中有二画键5个、三画键10个、三画以上键8个。它们是：

二画键5个：

丿	14→ S 键	十	23→ V 键
厂	24→ X 键	亻	43→ M 键
人	45→ I 键		

三画键10个：

氵	114→"5"键	一	121→A键
丁	231→R键	土	232→G键
艹	233→C键	大	245→T键
乚	321→E键	凵	323→"7"键
𠂉	421→P键	夕	424→N键

三画以上键8个：

讠	1231→"5"键	辶	124245→4键
扌	2314→B键	尸	2324→Q键
木	2341→W键	冂	3231→Y键
彐	2322→"9"键	钅	422231→Z键

此外，该方案还选了8个简码字，把它们直接标刻在键位上，使用时只需按键一次便可以输入。这8个字及其对应的键位是：

有→"2"键　　的→"8"键　　口→H键
是→"3"键　　我→"1"键　　女→"0"键
不→O键　　日→U键

由于这些字的使用频度高，有的还可以当作部首和常用的笔画组合，因此使该方案的动态码长显著缩短，提高了输入效率。

按照以上设计，人们在应用基本笔画笔顺编码法输入汉字时，除使用基本键外，若能自如地运用多画键，则每个汉字的平均按键次数为3.67次，高于英文平均每字5次的输入效率。

基本笔画笔顺编码法的键盘键位排列，参阅下节图3.12《基本笔画笔顺编码法36个基本键和多画键的键盘布局》。

2. 五笔画汉字输入法

五笔画汉字输入法是我国优秀的中年科技工作者、河南王永民设计的简易型汉字输入法，也是一种简单易学的汉字基本笔画编码输入方案。

该方案按照我国汉字书法的惯例，以“一（横）、丨（竖）、丿（撇）、㇏（捺）、乙（折）”五种单笔笔画作为构成汉字字形的基本笔画，并依次用“1、2、3、4、5”五个数字来表示。输入汉字的规则非常简单：只需按照汉字的标准楷书字形，以习惯书写的笔画顺序，一笔一画地用数字代码按键输入。规定每个汉字不论笔画多少，最多取五笔。就是第一、二、三、四笔和最后一笔（末笔）。如果一个字的笔画数不够五笔，则以数字键“0”结束；如果刚好五笔，则不必用结束键。例如：

“总”取丶丿丨乙㇏（第1，2，3，4笔和末笔）→

“43254”

“数”取丶丿一丨㇏（第1，2，3，4笔和末笔）→

“43124”

“用”取丿乙一一丨（正好五笔）→“35112”

“汉”取㇏㇏一乙㇏（正好五笔）→“44154”

“书”取乙乙丨丶 （不足五笔补0）→“5 5 2 4 0”

“大”取一丿丶 （不足五笔补0）→“1 3 4 0”

该方案设置了一个学习键“6”。在输入时遇到一时难以确定的笔画时，可以用数字键“6”来代替未确定的代码。例如“花”字，如果不知道最后一笔是撇（丿）还是折（乚），则可以输入“1 22 3 6”，最后一个“6”键代替未知的末笔，然后从显示屏中挑选出所需要的字。

从举例中可以看出：按惯例，凡是“点”笔都归入“捺”笔，“提”笔都归入“横”笔。其理由已在第2·3节中谈到了。“折”笔包括的范围比较广，除了“横、竖、撇、捺”以外的所有带折笔画，都归入“折”笔。

五笔画汉字输入法与基本笔画笔顺输入法一样，非常简单易学。只要会写汉字的人，十几分钟即可学会输入汉字。但是，由于重码字数量很大，几乎每次输入都要进行“二次选择”，汉字输入效率很低。例如输入“1 2 3 4 1”以后，机器将会显示出“本、相、李、格、检、植、柏”等一大串重码字，使用很不方便。因此，常把它作为王永民汉字编码体系（简称“王码体系”）中另外两种输入编码方案——五笔字型和五笔桥的辅助输入手段，适用于仅仅输入极少量汉字的场合。

3. 汉字笔形编码法

《汉字笔形编码法》是北京师范大学教授李金凯等研究设计、并最早取得国外专利的汉字字形编码法。该法的编码规则也比较简单，而汉字输入效率比上述两种基本笔画输入法大大提高，是一种具有部件编码特点的基本笔画编码输入方案。

该方案把构成汉字字形的基本笔画规定为 8 种，其中单笔笔画 6 种，复笔笔画 2 种。就是在“横、竖、撇、捺、折”的五笔画基础上，把“折笔”分解为“弯””笔（即“左折”）和“拐”笔（即“右折”）两种；又增加了两种复笔“叉”（“十”）和“方”（“□”）。并以“0～7”八位数字作为这 8 种笔画的代码。实践证明，把基本笔画分类为 8 种，可以有效地减少笔画编码方案中的重码现象，提高汉字输入的操作效率；而且八位数正好与计算机的计数方式相适应，便于处理。汉字笔形编码法的笔画名称、特征及代码如表3·1所示。

表3·1　　笔形编码法代码表

代码	1	2	3	4	5	6	7	0
笔划	一	丨	丿	丶	㇇	ㄴ	十	口
	㇀	丿	ノ	㇏	亅	𠄌	×	口
名称	横	竖	撇	点	弯	拐	叉	方
特征	→	↓	ノ	㇏	㇁	㇄	+	□

该方案对折笔的取码规则是：在区分“左折”和“右折”笔形以后，对多折笔形的取码规定是，先“弯”后“拐”的作为“弯”笔取码。如“乙”字取“5”码，不取“6”码；先“拐”后“弯”的作为“拐”笔取码。如“专”字的第三笔取“6”码，不取“5”码。

代码表中的“叉”笔，是指所有“横竖交叉”和“撇、捺交叉”的复笔。不要把复笔分开取码，如“草”字，应取“77 0 17”（+ +□－+）；“文”字应取“417”（丶一×）；“乘”字应取“3751134”（ノ+㇆一一ノ㇏）。

代码表中的“□”笔，是指所有方形闭合的复笔，不要拆开。如“中”字应取“20”（1□）；“电”字应取“601”（乚□一）；“事”字应取“510511”（丿一□冂一一），等等。

从以上所举的字例中看出，笔形编码法的一个特点是在取码时不是按照书写的笔顺，而是根据笔形在下笔时的位置高低、左右为次序的。排序的规则是先高后低，先左后右。例如“中”字，按书写笔顺是先“方”后“竖”，按下笔时位置是先“竖”后“方”，应按后者取码，故为“20”，不是“02”。这种取码方法的好处是在照底稿输入时比较直观方便，并避免了由于书写时笔顺不一致带来的编码二义性。

该方案的另一个特点是对合体字采取先拆分部件，再对部件按笔画编码的方法取码。拆分部件的规则是：将具有明显分隔间距的笔画组合作为部件拆开，分别取码（但离散型笔画结构不拆分）。每个字最多取3个部件，按照先上后下、先左后右、先外后内的次序排列。对拆分部件超过3个的字，取第1、2部件和最后一个部件（末部件）。对部件选取笔画数的规定是：每个部件最多取前三笔，末部件有时可只取第一笔。例如“脑”字，先分为“月”、“㐄”两个部件，“月”取前三笔为“351”，“㐄”取前两笔为“41”，故“脑”的笔形码为“35141”。

采用“笔形编码法”，每个汉字平均击键3.5次，一般打字员每分钟可输入40～60个汉字。为了加快输入速度，该方案还设计了简码和词汇码输入。该方案的简码字及代码为：

代　码	0	1	2	3	4	5	6	7
简码字	口	一	是	的	这	了	乙	十

当需输入简码字时，只需击键一次便可。据统计，这8个简码字的使用频度约占9％，因此可以有效提高输入速度。

词汇码又称为词组输入，是利用机器的联想能力来提高输入速度的输入法。该方案的词汇码输入方法，是在输入某一单字以后，再按一下“词组”键，屏幕就显示出以该字为开头的若干词组，以词组的使用频度高低排列。使用频度最高的词组排在第一位，按一下“回车”键便可输入；排在第2位以后的词组，则按一下相应的数字键便可输入。例如，在输入“中”的笔形码20以后，按“词组”键，屏幕显示：“1.中国 2.中型 3.中止 4.中华……”等8个词组，再按回车键，“中国”一词已被输入。在使用词组码以后，平均按键4～6次，便可输入一个词组，使输入速度明显提高。

4. 五笔字型汉字输入法

《五笔字型汉字输入法》原称为《二十六键五笔字型（WBZX）汉字编码方案》，简称为《五笔字型》法，是河南王永民等经过多年研究设计而成，在实践中不断改进优化，因而得到广泛应用的一种汉字部件编码输入方案。

该方案在对汉字字形进行大量分析统计的基础上，把字形结构归纳成五种基本笔画和三种基本字型，并分别赋以数字代码，如表3·2和表3·3所示。

该方案还根据对汉字部件构字频度的统计，选取实用构字频度比较高的构字部件130种，称为基本字根。按照基本字根的首笔笔形代码归纳分类，分成5个区，每区又分成5个位。这样，将130种基本字根分成5×5=25类，分配到标准键盘的25个字母键键位上（“Z”键键位不占用），每个键位以区位号11，12，13……53，54，55共25组代码表示。每组代码是

表3·2　　五种基本笔画代码表

名称	代码	走向	笔　画　及　其　变　形
横	1	→	一 ㇀（略带上提）
竖	2	↓	丨 丿（略带左撇）亅（带钩）
撇	3	↙	丿 丿
捺	4	↘	㇏ 丶
折	5	转折	乙 ㇇ ㇌ ㇄ ㇉

表3·3　　三种基本字型代码表

名　　称	代码	图　　式	字　　例
左右型	1	⿰ ⿰ ⿲	的　结　湘
上下型	2	⿱ ⿱ ⿳	男　翌　曼
杂合型*	3	⿴ ⿺ ⿷ ⊠	困旭区重

* 杂合型是指除左右型、上下型以外的各种字型，包括包围型（四面包围、三面包围）和独体型。

一个两位数，占一个键位。键位的分配是根据基本字根的构字能力大小和操作者手指击键动作的合理化原则，将构字能力大的字根分配在中排的食指操作位置，然后向四面扩散。一般每个键位上有 2 ～10种字根不等。并从中挑选出一个有代表性的字根（多为成字字根）作为键名，以便于称谓和记忆。

所谓“成字字根”，就是能够作为单字用的字根或部件。五笔字型方案的25个键位键名字根区位号分配表，如表 3·4 所示。包括130种基本字根的键位分配总表，如图3·3所示。1988年 5 月重新公布的4·3版标准键帽键位图参看图3·14。

表3·4　　键名字根区位号表

区＼位 / 键名	1	2	3	4	5	分类特征
1	王	土	大	木	工	横笔起笔类
2	目	日	口	田	山	竖笔起笔类
3	禾	白	月	人	金	撇笔起笔类
4	言	立	水	火	之	点笔起笔类（水除外）
5	已	子	女	又	乡	折笔起笔类

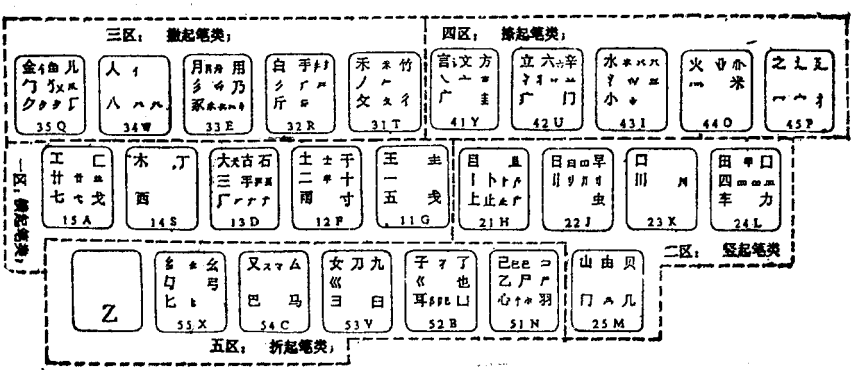


图 3·3　五笔字型字根键位分配总表

五笔字型方案的编码规则，是在基本熟悉了以上图、表所示内容的基础上，按照以下三条基本规则将汉字字形拆分编码。

（1）以书写顺序，从左至右，从上到下，由外到内，将汉字拆分成总表上已有的基本字根。

（2）取第1、2、3、末字根所在的键位（区位号或字母）作为汉字编码，每字最多取4码，不足4码时追加识别码。

（3）单体结构字形拆分时，要“取大优先”。就是在既可取笔画少的、又可取笔画多的字根时，优先提取笔画多的字根。

所谓“识别码”，是五笔字型方案为字形结构简单、字根数不足4个的汉字专门设置的标志码，用以区别重码字。识别码是一个两位数，一般是用这个字的末笔笔形代码和字型代码组合而成，举例如下：

字　例	末笔代码	字型代码	识别码
江	1	1（左右型）	11(G)
笔	5	2（上下型）	52(B)
冈	4	3（杂合型）	43(I)
必	3	3（杂合型）	33(E)

这里要说明两点：一是包围型的汉字末笔，规定取被包围笔画结构的末笔，不取外围结构的末笔。如“国”字末笔取“4”不取“1”，“冈”字末笔取“4”不取“5”，“远”字末笔取“5”不取“4”等等。二是对一些容易产生不同笔顺的末笔，统一规定取向右下角伸得最远的笔画。如“仇”和“化”字，都是取“5”不取“3”。

五笔字型也设置了简码和词汇码，以提高汉字输入速度。

该方案的简码分为三级。一级简码是25个高频汉字，分配在25个字母键键位上，它们是：

一(G)　地(F)　在(D)　要(S)　工(A)

上(H)　是(J)　中(K)　国(L)　同(M)

和(T)　的(R)　有(E)　人(W)　我(Q)
主(Y)　产(U)　不(I)　为(O)　这(P)
民(N)　了(B)　发(V)　以(C)　经(X)

一级简码的输入方法很简单，只需按一下该字所在键位，再加一下间隔键便可。注意把简码字与键名字(字根)区别开。若要输入表3·4所示的键名字，则需连续按此键4次才能输入。

二级简码是由汉字拆分后取前两个字根组成的代码，共有625个；三级简码是由汉字拆分后取前三个字根组成的代码，共有4400多个。由于记忆量大，一般很少使用。

该方案的词汇码，分二字词、三字词、四字词和多字词数种。输入词组时，不论长短一律取4个字根组成词汇码。

双字词，分别取两个单字的前两个字根组成词汇码。例如“经济”一词，取“纟又氵文”四个字根，词汇码为“XCIY”。

三字词，前两个单字各取其第一字根、第三个单字取其第一、二字根组成词汇码。例如“计算机”一词，取“讠竹木几”，词汇码为“YTSM”。

四字词，分别取每个单字的第一字根组成词汇码。例如“科学技术”一词，取“禾⺍扌木”四个字根，词汇码为“TIRS”。

多字词，取第一、二、三个字和最后一个字的第一字根组成词汇码。例如“中华人民共和国”一词，取“口亻人口”四个字根，词汇码为“KWWK”。

由此可见，用词汇码可以大大提高汉字输入速度。但是，设置词汇少了不敷应用，多了占用内存量太大。为了解决这一矛盾，该方案还设有用户自行造词功能。用户自造词组存入内存后，同样可用五笔字型码调用。

该方案还设置字母“Z”键为“选择式易学输入键”，简

称“学习键”。该键的作用是充当一切“未知数”字根的代表。当操作者在拆分字形时难以确定的字根，都可用“Z”键代替。例如输入“敬”字时，记不清第二个字根该怎么取，就可输入“廿Z口文”，若没有重码，“敬”字就可输入；若有重码，机器便将重码字全部显示出来，供操作者进行二次选择。输入一个单字时，使用“Z”键的次数不受限制。显然，“Z”键愈多，则重码字也愈多。如果输入四个“Z”，则机器显示全部几千个汉字供选择，就没有意义了。

《五笔字型汉字输入法》是一种高效型的汉字形码（部件）输入法，适合于年轻的专职操作人员使用。经过指法训练，并熟练掌握规则以后，最高可达每分钟输入160多个汉字的速度。但是，由于取码规则比较复杂，学习记忆量较大，非专职的人员学习掌握起来有一定的困难。为了解决这一困难，作者在1988年设计了一种介于“简易型”五笔画输入方案和“高效型”五笔型输入方案之间的“普及型”汉字输入方案。它是联接二者的“桥”，故定名为《五笔桥汉字输入法》。

5. 五笔桥汉字输入法

《五笔桥汉字输入法》是在《五笔字型汉字输入法》的基础上简化而成的。它是一种比较简单易学、又有较高的输入效率、适合于非专职的中年操作人员使用的汉字输入方案。

《五笔桥》方案的字根键位分配总表如图3·4所示。由图可知，该方案的键盘比《五笔字型》方案的字根键盘简单明了。它共有首根60种，单笔与其他复合笔画16个，在键盘上按笔画类型分作5个区，每区5个位，共25键。这些字根和笔画所在的键位与《五笔字型》的键位完全相同。

先介绍一下“首根”和“代表根”的概念。

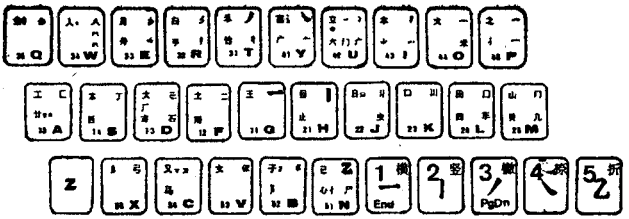


图 3·4 《五笔桥》字根键位分配总表

所谓“首根”，是指由首笔参与构成、且与汉字的剩余部分相分离的键面字根。所谓“代表根”，是指为实现从《五笔桥》向《五笔字型》过渡，保持键面字根的一致性，并尽量减少重码而设计的字根。《五笔桥》规定了以下代表根：

“礻”代表“礻”和“衤”；“丹”代表“舟”；“酉”按“西、一”输入；“鱼”按“鱼、一”输入；“足”按“口、止”输入。

以下介绍《五笔桥》方案的输入规则。

（1）键外字输入时，先打首根，再打剩余部分的第1、2、3笔和末笔的单笔画，共4码。不足4码时，打一下间隔键。如“根”字，打“木、乙、一、一、乀”；“仁”字打“亻、一、一、间隔键”。

无首根的字，按照《五笔画输入法》的规则，用单笔画输入。单笔画“横、竖、撇、捺、折”分别安排在“G、H、T、Y、N”五个键上。例如“重”字，打“TGHNH”（丿、一、丨、乙、一）；“本”字打“GHTYG”（一、丨、丿、乀、一）。

（2）键面字（即键面上标有的字）输入时，按照《五笔画输入法》依书写顺序拆成单笔画，取1、2、3、4笔和末笔的单笔画输入。如不够5码时，打间隔键结束。例如“石”

字，打“GTHNG”（一、丿、丨、乙、一）；“工”字打“GHG间隔键”（一、丨、一）。

对于键面左上角的第一个字即键名字的输入，除按照以上方法外，还可以按照《五笔字型》的方法，“连打四下”输入。

对于“G、H、T、Y、N”五个键右上角的单笔画输入，则把所在键“连打五下”。例如“乙”，打“NNNNN”便可输入。

（3）高频字输入。《五笔桥》的高频字设置以及输入方法与《五笔字型》的一级高频字完全相同。

（4）选字与联想输入。当按照以上规则（1）的方法输入代码后，所需输入的字出现在提示行中时，可用键盘最上一排的数字键依其序号挑选。一旦按下该字的序号数，与该字可构成常用词组的字便会自动依次显示在提示行中，等待第二次挑选。这叫做“联想输入”。用联想输入能显著提高输入速度。

（5）默认第一个。当按照要求输入一个字的代码后，如果所需的字出现在提示行的第一个位置上，这时不必再按序号挑选，也不用打“回车”或“间隔”键，只管继续输入下文，这个字便自动输入。这叫做“默认第一个”。如果要输入的字在第一个位置，而下文是一个标点符号，也可以照样“默认”输入。

（6）学习键“Z”键的作用与《五笔字型输入法》相同。

6. 联合45-3汉字拼形输入法

《联合45-3汉字拼形输入法》是上海876所等七个单位通力合作，有虞柏年等参加，在原联合攻关的编码输入方案——《振兴1号汉字拼形输入法》的基础上，经过计算机辅助设计

优化而成的汉字部件编码输入方案。

该编码方案设计的汉字容量为国标GB2312-80基本集的汉字6763个。选取构成这些汉字字形的主部件共170个。以起笔笔形分区，将主部件分成“横、竖、撇、点、折”五区，并按照部件实用频度高低和手指操作合理负担的原则，将它们分配在标准键盘的45个键位上，每个键位上一般为4个部件，看上去比较清楚醒目。部件键位图如图3·5所示。在单个汉字输入

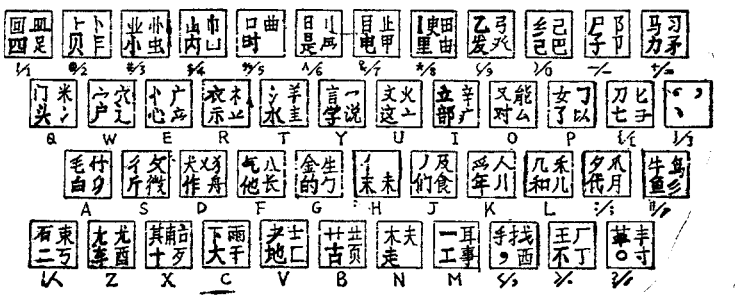


图 3·5 《联合45-3拼形输入法》部件键位图

时，在45键范围内，每字击键最多3次，故称为“45-3拼形输入法”。据统计，该方案的平均动态码长为2.67，动态重码率0.21%。其中一级汉字重码46对，二级汉字重码109对。经过学习和指法训练并熟练操作后，能达到较高的汉字输入速度。

以下简要介绍《联合45-3汉字拼形输入法》的编码规则。

首先要基本熟悉规定的170个部件和它们的键位分布图(图3·5)。输入时，按照每个汉字的形体结构和书写笔画顺序，把字形拆分为已规定的部件（包括单笔部件），然后依次取首、次、尾三个部件组成该字的输入码组，击键输入。例如“赣”字，取首、次、尾部件是：立、日、贝，找到它们所在

的键位Ⅰ、6、2，即可击键输入。又如“荷”字，取“艹、亻、口”三个部件，击打“B、H、5”三键输入。又如“五”字是单体字，就取“一、丨、一”三个单笔部件，击打“M、8、M”三键输入。总之，在正确拆分的基础上，只取首、次、尾三个部件，其它的部件或笔画均可省略。

对于只有两个部件组成的汉字输入时，需要区分一下字型结构。该方案对字型区分要求比较简单，只需区分左右型和其它（杂合）型两种。对左右型结构的汉字，在输入两个部件的代码后，以“间隔”键结束；对除左右型以外的其它型结构的汉字，则在输入两个部件代码后以“Y”键结束。由于汉字的两部件结构的字形较多，这样规定以后，可以大大减少重码字。现举例说明：

汉　字	字　型	部　件	输　入　码
洋	左　右	氵　羊	TT间隔键
取	左　右	耳　又	MO间隔键
李	其　它	木　子	N－Y
风	其　它	几　乂	LDY
司	其　它	冂　㠯※	PXY

※在有的编码方案中，没有设“㠯”部件时，就需拆分为“一、口”两部件。

该方案为了解决在拆分部件时可能产生的模棱两可的情况，规定了“从大取码”和“从整取码”的原则。所谓“从大取码”，就是在字形分解时要按笔顺抽取最大部件，又称为“大元优先”。所谓“从整取码”，就是在字形分解取码时要

照顾相邻部件的完整性，不要在抽取最大部件时把下一个部件拆散。对独立的交笔部件，应把它看作一个整体，不予拆分。现举若干字例说明，如表3·5所示。

表3·5　　部件分解和取码举例

字例	首、次、尾部件		取码原则
	应取	不取	
辨	辛、辛	立十十	从大取码，有“辛”部件
高	古冂口	亠口口	从大取码，有“古”部件
尊	丷酉寸	兰门寸	照顾第二部件的完整
戥	贝一戈	贝二	照顾第二部件的完整
卷	丷夫巳	兰大巳	“夫”、“未”是独立的交笔部件，不拆分。
朱	丿　未	𠂉　木	

对于少量汉字在书写习惯上有多种笔顺而使部件分解时可能产生的麻烦（即一字多码现象），该方案设计了相应的容错码，可以容许对同一个汉字字形有二至三种不同的部件分解编码均能输入。例如“义”字，部件分解时可以容许先“、”后“乂”或者先“乂”后“、”；“脊”字分解时，可以容许以“丬、人、月”为序，或者以“人丬月”为序输入。又如常用的“办”字，可以分解为“力丶丶”输入，也可分解为“力𠂉”（两部件）输入。这就进一步方便了使用。

《45—3拼形输入法》也设计了简码字和词汇码输入。根据汉字频度统计，该方案还以键面左下角的43个高频字和3个常用标点符号作为简码字，输入时只需按一下该字所在键及“间隔”键即可。该方案的词汇码输入，只需取3个部件。即

该语词的第一、二个汉字的起始部件和最后一个汉字的末部件作为语词的首、次、尾部件。取码方法举例如表3·6所示。

表3·6 语词取码方法举例

语　　词	首、次、尾部件
社会主义	礻　人　义
四个现代化	四　个　匕
现代汉语词典	王　亻　八
中华人民共和国	口　亻　、

7. HPX汉字拼形输入法

《HPX汉字拼形输入法》是北京李慧勤等研究设计的汉字部件编码输入方案。

该方案以国标GB2312-80中的6763个汉字和682个图形字符作为编码汉字图形字符集，选用基本构字部件104个，按照起笔笔形分成“点、横、竖、撇、折”五区，将它们合理分配

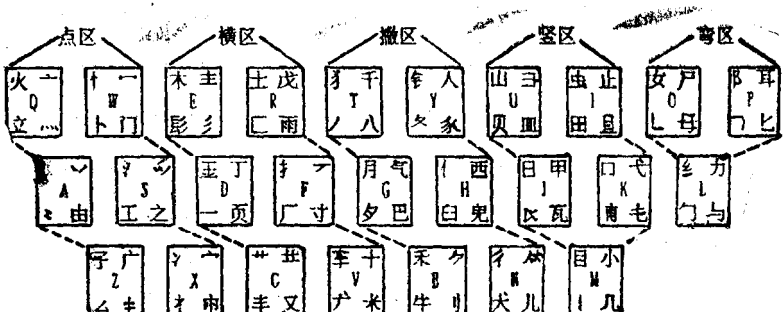


图 3·6 基本部件键元分布表

在标准键盘的26个字母键键位上，每个键位上设有4个基本部件，称为键元。键元的键位分配如图3·6《基本部件键元分布表》所示。

除了键元分布表上的104个基本部件外，该方案还设置了与基本部件相应的扩展部件，如下表所示：

基本部件	扩展部件	基本部件	扩展部件
忄	心 ⺗	月	月
卜	卜	日	曰
戈	戋 戉	纟	幺
匚	匚	勹	勹
千	干	丰	夫
山	屮	艹	廾
彐	彐	丰	彐 ⺕
阝	卩	牛	牛 牜
耳	耳	刂	刀 刂 彡
匕	七 匕	丨	亅
讠	言	几	凡
氵	水 氺	、	乀
之	辶 廴	丁	丁
扌	手 ⺘	厂	厂

该方案的汉字字形拆分编码规则比较简单易学，遵循习惯书写顺序，将汉字字形分解成键元表上已有的部件，然后选取第1、2、3、末部件，按键输入。汉字输入编码的最大码长

为4，对不足4码的汉字输入，以字型区分。字型也只需区分“左右型”和“其它型”两类。左右型结构的字型用“H”键结束；其它型结构的字型用“间隔”键结束。左右型结构的字用“H”键以后仍不足四键时，再加间隔键结束。

对字形结构紧密的单体字，拆分部件的原则是抓住汉字字形嵌套结构的核心，笔画不断，先抽取首笔形成的最大部件作为第1部件，然后依次取第2、3、末部件。举例如下：

单体字	首部件	2、3、末部件
重	千	曰、一、一
束	十	口、八
平	干	丿

后两字拆分后均不足四键，需加“间隔键”结束。

对占汉字多数的合体字，都能够比较形象、直观地按部件拆分字形。但需注意，在拆分过程中对键元表上已有的部件不再拆分；拆分后应以书写习惯顺序排列编码输入；每字四码，不足四码时根据字型区分打不同的结束码。举例如下：

合体字	字型	拆分部件	结束键
好	左右	女、子	H、间隔键
利	左右	禾、刂	H、间隔键
获	其它	艹、犭、犬	间隔键
得	左右	彳、日、一、寸	（不加）

除了上述一般规则外，还需补充说明以下几点：

（1）对左右型的多个部件字形，在拆分时先取左边部分的第1、2部件，再取右边部分的第1、末部件。例如“赣”字，可拆分成6个部件，则应取“立、日、攵、贝”4个。

（2）如果字形拆分时末部件出现在包围型笔画结构（如“□、冂、匚”）里面，则只取外框，不取内部。例如“蒿”字，可拆分为“艹、亠、口、冂、口”5个部件，末部件应取“冂”，不取“口”。

（3）键元字的输入。键元表（图3·6）上的单字称为键元字。其中，在左上角的为基本键元字，其它为扩展键元字。它们的输入格式是：

基本键元字：键元字所在键位＋尾笔＋“H”＋“间隔键”

例如：“木”为“EAH间隔键”

扩展键元字：键元字所在键位＋尾笔＋“H”＋“H”

例如：“气”为“GOHH”

（4）高频字的输入。该方案设置高频字24个，常用标点符号2个。输入时只需按一下高频字所在键位和间隔键即可。高频字在26个字母键键位的分配如下（按标准键盘从上到下、从左到右为序排列）：

以（Q）　为（W）　一（E）　地（R）　他（T）
人（Y）　有（U）　工（I）　我（O）　动（P）
个（A）　是（S）　的（D）　发（F）　国（G）
和（H）　，（J）　。（K）　了（L）
上（Z）　下（X）　到（C）　在（V）　不（B）
大（N）　们（M）

（5）智能键的使用。该方案设置“？”键作为“智能键”，供操作者在不太熟悉键元表时使用。其功能与《五笔字

型》方案中的“Z”键一样，可以用来替代任何部件。但是规定在输入一个汉字时，“？”键只能使用一次。例如“高”字，若不知“亠”在那个键位上，可用“？KUK”输入。

（6）词汇码输入。该方案的词组输入方式，分为二字词输入和多字词输入两种，分别使用词汇输入键“；”键和“·”键。

在二字词输入时，先打第一个汉字的前两个部件代码，然后打“；”键，紧接着再打第二个汉字的前两个部件代码。例如“答复”的词汇码是“NY；GJ”；“高度”的词汇码是“GK；ZC”。

在多字词（三字词或四字词）输入时，先打“·”键，然后需输入词汇中每个单字的声母。如果是三字词，打“间隔”键结束；四字词则自动结束，举例如下：

三字词：“计算机”的输入码是“·JSJ间隔键”。四字词：“联系实际”的输入码是“·LXSJ”；“满怀信心”的输入码是“·MHXX”。

该方案的编码规则比较简明易学，汉字输入速度较高，曾在全国输入编码方案评测中获得好评。

8.仓颉字母输入法

《仓颉字母输入法》又称为《天龙仓颉造字法》，简称《天龙法》。它是台湾朱邦复等经过长期研究设计并较早获得实际应用的一种汉字繁体字形编码输入方案。

该方案以24个称为“仓颉字母”（或“中文字母”）的汉字，作为24种基本部件和笔画的称谓，连同由它们引伸演变而来的60个辅助字形（也是部件或笔画）去构成全部通用汉字约一万六千多个繁体字形。用标准键盘输入，每个汉字不超过五

码，单体字不超过四码。该方案规则严谨，重码率很低，但学习记忆需要下一番功夫。

为了便于记忆，该方案把24个“仓颉字母”分成四大类，即哲理类、笔画类、人体类、字形类，分配在字母键除X键和Z键外的各个键位上。如表3·7所示。

表3·7　　“仓颉字母”分类表

哲理类		笔画类		人体类		字形类	
日	A	竹	H	人	O	尸	S
月	B	戈	I	心	P	廿	T
金	C	十	J	手	Q	山	U
木	D	大	K	口	R	女	V
水	E	中	L			田	W
火	F	一	M			卜	Y
土	G	弓	N	难字	X	重字	Z

表中哲理类和人体类11个“仓颉字母”，就是汉字的常用成字部首，比较直观明了。笔画类和字形类13个“仓颉字母”代表不成字的基本笔画和部件，如表3·8所示。为了便于记忆，表中对每个笔画和部件都规定了读音。

上述24个仓颉字母，除代表基本笔画及部件外，还代表由它们引伸而来的60个辅助字形，如表3·9所示。

该方案的编码规则，是在熟悉了以上的“仓颉字母”及其辅助字形的基础上，将汉字字形先分解成“字首”和“字身”两部分。字首部分取一至二码，字身部分取一至三码，然后按照“由外而内，由上而下，由左而右”的顺序取码。

表3·8　笔画类和字形类仓颉字母表

笔画类	仓颉字母	竹	戈	十	大	中	一	弓
	代表笔画	丿	丶	十	ㄨ	丨	一	亅
	读　音	斜	点	交	叉	纵	横	勾
字形类	仓颉字母	尸	廿	山	女	田	卜	
	代表字母	コ	艹	凵	女	口	卜	
	读　音	侧	并	仰	纽	方	卜	

表3·9　仓颉字母辅助字形表

仓颉字母	辅助字形	仓颉字母	辅助字形
日	曰	手	龵 丰 牛 扌
月	冂 冖 ⺈ 𠂊	水	又　氵
金	丷 八 儿	火	小 忄 灬
木	寸 ヤ	土	士
戈	丶 厶 广	竹	丿　厂
十	亠	口	（无）
大	𠂇 ㄨ	尸	彐 匚 ㇆ ⺕
中	肀　丨	廿	艹 龷 䒑
一	工　厂	山	凵 屮 ㇄ 乚
弓	亅 乛 ㇇	女	く 乚 ⺈
人	く 乀 𠂉	田	口 ⺈
心	匕 七 勹 忄 ⺗	卜	⺊ 亠 二 丫

对于字体本身笔画相互交连的单体字，就以抽取首笔形成的第一码（笔画或部件）为字首，所以称为“首码必取”。除了字首以外剩余的部分都称为字身，字身至多取三码。字身不足三码者全取，超过三码者取首、次、尾三码。也就是说，单体字总共取码不超过四码。现举若干字例说明。

汉字	字首	字　身	取　码
妻	十	弓　女	十中女
商	亠	丷冂八口	卜金月口
雨	一	丨冂丶丶	一中月卜
函	了	凵火	弓山水
舟	丿	丿亠丶	竹月卜戈

对于字体可以左右或上下分离的合体字，取其最上方或最左侧的字形为字首，其余部分为字身。字首的取码规则是，若一码取足，就直接取此码。如“枝”字取“木”，“们”字取“人”，“鹏”字取“月”，“厚”字取“厂”等等，均只取一码。若一码不能取足，则取二码。超过二码时，限取首尾码。例如：

汉字	字首	取码	汉字	字首	取码
爬	爪	竹人	惑	或	戈一
牌	片	中中	威	戊	戈竹
建	廴	弓大	疫	疒	戈卜

字身的取码规则与单体字相同，可取一至三码。如果字身又可上下或左右分离者，在左侧或上方的字形称为“次字

首”，其余部分称为“次字身”。其取码规则是：次字首能够一码取足时，次字身取首、尾二码；次字首不能一码取足时，则取次字首的首尾码及次字身的字尾码。现举字例说明。

汉字	字首	字身	次字首	次字身	取码	说明
字	宀	子			十弓木	字首一码取足
春	夫	日			手人日	字首取首尾码
睡	目	垂			月山竹十一	字身取首次尾码
彬	木	彡	木	彡	木木竹竹	次字首一码取足
淹	氵	奄	大	电	水大日山	次字首一码取足
最	日	取	耳	又	日尸十水	次字首取首尾码
解	角	刀	刀	牛	弓手尸竹手	次字身取字尾码
辨	辛	辛	辛	辛	卜十戈中十	字首取首尾码，次字首取两码，次字身取字尾码

仓颉编码法取码时，还应遵守以下三条原则：

（1）精简原则。当一字有多种分解的可能性时，以取码

最少的分解法为正确。如“王”字取“一土”，不取“一十一”；“乍”字取“竹尸”，不取“人卜卜”等。

（2）完整原则。也就是“大元优先”或“从整取码”。当一字有多种分解法，且取码数相等时，应依顺序取字形较复杂的码，再取字形较简单的码。如“丰”字应取“手十”，不取“十手”；“民”字取“口山心”，不取“尸山心”等。

（3）照顾字形特征的原则。在符合“精简”和“完整”两原则的前提下，取码时应照顾到能最贴切地表示字形特征。例如“之”字应取“戈弓人”，不取“卜竹人”（第二笔不切断）；“者”字应取“十大日”，不取“土竹日”（交叉笔画不拆开）等。

该方案规定对组字时常用的11种字形，不论它们在字首或字身，都只取二码。这些字形称为复合字，它们是：

复合字	取码	复合字	取码
门	日弓	目	月山
佳	人土	气	人山
鬼	竹戈	戋	女戈
八	竹山	忄	卜心
阝	弓中	亠	卜口
冂	中弓		

该方案还设有“难键”（“X”键）和“重键”（“Z”键）。难键相当于“智能键”，在分解字形中遇到困难，不知如何取码时，可以用此键代替。难键的使用规定是：当首、尾码易取时，可取“首、难、尾”，例如“身”字，不知中间一

码怎样取，可以打“竹、难、竹”三码输入。当尾码也不易取时，可以只取“首、难”二码。例如“臼”字，只知首码为“丿”，可以打“竹、难”二码输入。

“重键”（“Z”键）是用来处理重码字输入的。仓颉编码法的重码率比较低，在一万六千多个汉字中仅有105对，每对只2字。在遇到重码字时，机器同时显示两字，若需要输入的字排在后面，则需按“重键”，便能将该字输入。

9. 大众汉字拼形输入技术

《大众汉字拼形输入技术》，简称《大众》码，是四川陈代于、刘明华等研究设计的汉字部件编码输入方案。

该方案以国标基本集（GB2312-80）中的一、二级汉字作为编码字集，并留有扩充余地。选择了172个笔画、笔画组合、偏旁部首作为字元（即部件）。其中有基本字元128个，联想字元44个，采用“字元归并优化算法”程序进行计算机辅助设计，合理地把172个字元归并成26组。在仅采用24个高频简码字的情况下，使单字的重码实用概率总和降至0.378%。平均击键次数（含间隔键一次）为3.42键/字。在单字输入方式下，熟练操作员每分钟可输入120个汉字。采用词汇输入时还可达到更高的速度，曾在全国性的汉字输入方案评测中获得好评。

该方案的一个最大特点是：为了帮助记忆字元及其键位分布，创造了一种“造字记忆法”。就是把对应于同一个键位上的所有基本字元和尽量多的联想字元进行适当组合，构成一个生造汉字。对应26个字母键键位，共有26个生造汉字。例如在字母“K”键上安排了五个字元：“山、言、一、尚、业”，经组合构成了一个生造汉字“䜛”。这26个生造汉字的键位布

局如图3·7所示。

图 3·7 大众汉字拼形编码方案键盘布局图

为了便于使用，该方案还给每个生造汉字赋予一个读音。就是选择一个与生造汉字形状相似，又与所在键位有联系的真正汉字作为生造汉字的读音，称为“读音字”。例如生造汉字“鼍”的读音是“凯”，操作者即可据此判断出构成这生造汉字的5个字元都在“K”键上。实验证明，用生造汉字的办法把本来互不相关的多个字元结合成一个整体，能有效地帮助对一百多个字元及其归并分类情况的记忆。人们只需会写会认26个生造汉字，就等于掌握了编码方案所要求的主要记忆内容，这比死记硬背一堆零散字元要容易得多。

该方案的字形拆分编码规则比较简单易学。就是在记住26个生造汉字的基础上，对所有汉字一律按书写顺序和大元优先准则取其第一、二、三、尾字元。如前所述，“大元优先”就是每次提取字元时应按书写顺序提取小于整字的最大可能字元，即笔画数尽可能多的字元。每字取4个字元，不足4个字元

时，不需要追加区分字型结构或笔画特征的码键。不但减少了操作人员的记忆量，并使平均击键次数明显降低，提高了汉字输入速度。例如："汉"字可分解为2个字元"氵"和"又"，它们分别在"C"键和"I"键上，故该字的大众码为"CI"；又如"码"字可分解为4个字元"丆口冂与"，它们分别在"I"键、"D"键、"E"键和"N"键上，故该字的大众码为"IDEN"。

为了进一步缩短平均动态码长，该方案还设置了24个高频字简码，输入时只需按一下该字所在键和"间隔"键便可。这24个简码为：

乙（A）、就（B）、到（C）、八（D）、为（E）、是（F）、国（G）、几（H）、不（I）、的（J）、一（K）、民（L）、在（M）、得（N）、这（O）、力（P）、经（Q）、日（R）、他（S）、部（T）、人（U）、地（V）、大（W）、我（X）、年（Y）、已（Z）。

10. 见字识码输入法

《见字识码输入法》是上海支秉彝博士研究设计的汉字音形编码输入方案。该方案的基本原则是建立在汉字字音和字形双重关系基础上。具体做法是：将一个汉字字形尽可能依照其拓扑结构拆分成4个字元（部件或笔画）组成的字元串，再以它们的读音（或者它们的"关系字"的读音）第一个字母组成代码组，用标准键盘输入汉字。为了便于说明，举一个典型的字例：如汉字"路"，可以拆分成"口、止、文、口"4个字元，它们读音的第一个字母分别是"K、Z、W、K"，这就是"路"字的输入代码组。

该方案规定的汉字输入代码组是等长码，4位一组。对于不是由4个字元组成的汉字，应遵循如下取码规则：

（1）对于由3个字元构成的汉字，则应将其最后一笔作为第4字元编码。例如“待”字，是由“彳、土、寸”3个字元构成，则还需加上最后一笔“、”，成为“彳、土、寸、、”4个字元。

（2）对于由2个字元构成的汉字，在拆分两个字元以后，除将最后一笔作为第3字元外，还须将该字本身作为第4字元。例如“打”字，拆分两字元是“扌、丁”，最后一笔是“亅”，应取的4个字元就是“扌、丁、亅、打”。

（3）对笔画数很少的汉字，作为字元时可不拆分，但作为单字时仍需拆分。例如“天”字，需拆分为“二、人”，再以两个部件汉字的规则取码。

（4）对多于4个字元的汉字，用“前拆后拼”的方法检取4个字元。例如“裹”字，可拆分为“亠口口于八”5个字元，再把两个“口”拼成一个“回”，则应取“亠回于八”4个字元。

（5）为了减少重码，该方案规定以下7个单字作为字元时不拆分。它们是：

金、页、可、鸟、皿、乍、舟

现将上述规则归纳举例如表3·10所示。

在确定了汉字拆分的字元串以后，“见字识码”方案的另一个关键是确定字元的“关系字”。确定关系字后，将它读音的第一个汉语拼音字母（称为“标记”）代入字元串中，就可得到该字的输入代码。

那末，怎样确定“关系字”呢？

如果字元本身是一个单字，则关系字就是它自己。例如上例中的“口”、“止”、“文”等，它们的标记就是“K、Z、

表 3·10　　“见字识码”取码规则简表

字元数	占常用字总数%	汉字	字元、关系字标记	末笔标记	字本身标记	字码
4	50%	路	口止文口 K Z W K			K Z W K
		寨	宀回 亍八 A H J B			A H J B
3	35%	余	人干八 R Y B	丶 D		R Y B D
		患	口中心 K Z X	丶 D		K Z X D
2	15%	吴	口　天 K　T	乀 N	吴 W	K T N W
		天	二　人 E　R	乀 N	天 T	E R N T

W”。如果字元不是单字，则应按以下 4 条规则选取关系字：

（1）字元是笔画时，以该笔画的名称作为关系字。例如“乀”的关系字是“捺”，标记为“N”；“丨”的关系字是“直”，标记为“Z”；“丿”的关系字是“撇”，标记为“P”等。

（2）字元是有称谓的偏旁部首时，以它们的习惯名称为关系字。例如“亻”的关系字是“人”，标记为“R”；“阝”的关系字是“耳”，标记为“R”；“犭”的关系字是“犬”，标记为“Q”；“虍”的关系字是“虎”，标记为“H”等。

（3）字元的外形相似时，都归成一类，以其中的一个常用字为关系字。例如“刂、丿、丶、㇀”等，都归入“丨”

类,关系字是"直",标记为"Z";又如"彐、彐、彑"等,归入"彐"类,关系字"雪",标记为"X";"⺶、⺷、⺸"等归入"羊"类,标记为"Y",等。

（4）少数没有关系字的字元,规定以"O"为标记。如"德"字的第2字元"𠁣",它的标记为"O"。

为了能大体平均使用26个字母,避免标记过分集中于某些字母的现象,以减少重码,该方案专门指定12个字母作为若干常用字元的标记,如表3·11所示。

表3·11　12个常用字元的标记表

序号	字母	标记的字元
1	A	讠亠子亠言
2	B	山
3	D	𠂊鱼魚
4	E	日曰
5	F	幺纟幺糸糸
6	H	乙𠃊乛丁乚㇉乚𠃌㇀く㇄乛
7	I	一衣𧘇𧘇𧘇衤衤覀
8	L	车彳車
9	O	手龵扌几隹
10	P	艹廾
11	U	水氺⺗氺氵㐅丿
12	V	木朩朩

“见字识码”方案的汉字输入码是等长码，码长为4码，打字输入时还需加一个“间隔”键，故需击键5次。为了提高汉字输入速度，该方案规定了由1～3个字母组成的144个汉字快速码。其中对26个使用频度最高的汉字规定了一字一键（再加“间隔”键）的简码，它们是（以字母顺序排列）：

地（A）、不（B）、产（C）、的（D）、而（E）、分（F）、个（G）、和（H）、以（I）、进（J）、可（K）、了（L）、们（M）、年（N）、口（O）、表（P）、起（Q）、人（R）、是（S）、他（T）、水（U）、这（V）、为（W）、学（X）、有（Y）、在（Z）

此外，该方案还规定了词汇码。它的输入规则是：选取组词单字读音的第一个拼音字母组成词汇码的代码组。二字词的码长为2码，例如“安排”为（AB），“今后”为（JH），“科技”为（KJ）等。三字词和四字词分别用3个和4个字母组成代码组。例如“无线电”为（WXD），“工程师”为（GCS），“无产阶级”为（WCJJ），“科学技术”为（KXJS）等等。该方案的词汇码比较好记好用，常用后可大大提高汉字的输入速度。

11. HYX汉字音形输入法

《HYX汉字音形输入法》是前述《HPX汉字拼形输入法》的作者为稍懂汉语拼音的人员设计的一种音形编码输入方案。这两个方案采用相同的部件分类和键元表（参见图3·6）。对汉字字形的拆分排序规则也完全相同。汉字码长为4码，不加“间隔”键，不足4码时则一律以间隔键结束。

该方案是在对汉字字形以部件拆分的基础上，首先选取该字的声母作为第1个码元，然后依次取第1、2、末部件作为

第2、3、4个码元，其代码格式为：

汉字声母＋第1部件＋第2部件＋第末部件

例如“获”字的输入码：先分解字形成“艹”（“C”键）、“犭”（“T”键）、“犬”（“N”键）三部件，再冠以该字的声母“H”，就是“HCTN”。

对字形拆分不足3部件的汉字，在冠以该字声母后，以间隔键结束，不必区分字型。例如“好”字，分解成两个部件“女”、“子”（在字母“O”、“Z”键上），声母是“H”，则该字的输入码是“HOZ”＋“间隔”键。

关于键元字的输入。该方案规定，凡在图3·6键元表上标有的单字统称为键元字。键元字的输入代码格式是：

键元字声母＋所在键位＋尾笔＋“间隔”键。例如键元字“子”，所在键位为“Z”，尾笔为“－”（“D”键），声母为“Z”，则输入码为“ZZD”＋“间隔”键。

该方案选用的高频简码字及其输入规则与前述《HPX汉字拼形输入法》相同。

词汇键“；”和“·”键的使用规则也与《HPX汉字拼形输入法》相同，只须注意：二字词输入时，每个汉字的两码为：该字的声母＋第一部件。

例如“答复”的词汇码是：“DN；FY”；“高度”的词汇码是“GQ；DZ”。为了方便使用，该方案规定：如果在二字词输入时，其中有一个字是高频字，则仍可按照高频字输入规则输入。例如“动”是高频字，（在“P”键），则“动作”的词汇码是“P间隔键；ZH”；“国”是高频字（在“G”键），则“美国”一词的词汇码是：“MA；G间隔键”。

三字词和四字词的输入方法，也与《HPX汉字拼形输入法》完全相同，兹不赘述。

12. 通信拼音字

《通信拼音字》，简称“通拼”，是总参通信部杜牧平在长期军事通信工作实践的基础上，经过多年研究设计的一种汉字拼音编码方案。该方案的主要特点是字码一一对应，无重码，汉字输入速度快。它主要用于军事通信和自动化指挥系统，经过短期训练的操作人员，能够达到每分钟一百多个汉字的输入速度。

该方案以国家颁布的《汉语拼音方案》为基础，用26个字母实现声韵双拼，再以音（调）义定字的方法进行编码。码集包括通用汉字8000多个。每个汉字一般用3个字母表示，其代码格式是：

声母＋韵母＋调义定字字母

该方案规定：对常用的470多个汉字，（它们的使用频度占70%以上）其代码只用“声母＋韵母”，不加定字字母，故动态平均码长为2.5。该方案码集中的汉字与代码对应性好，见字知码，见码知义，没有重码字及二义性编码。

该方案采用“双拼”法，即用两个字母分别表示声母和韵母来拼读一个汉字音节，其中前面一个字母是声母，后面一个字母是韵母。对于汉语拼音方案中“ZH、CH、SH”三个双字母声母，为了保持代码组长度的一致性，分别用三个单字母“V、W、Y”来替代，其替代关系如下：

汉语拼音方案	通信拼音字
ZH	V
CH	W
SH	Y

声母中的其它字母，用法与汉语拼音方案完全相同。

韵母的情况比较复杂。因为汉语的韵母有30多个，为了也能用26个拉丁字母的单字母表示，该方案规定有7个字母作为“一码两韵”。这是在研究分析汉语声韵母搭配规律的基础上，成功实现了使一个字母当两个韵母使用时不会产生混淆现象。现将汉语拼音方案（简称“汉拼”）与通信拼音字（简称“通拼”）的韵母替代关系列出如下：

汉拼	通拼	汉拼	通拼	汉拼	通拼
a	a	iong(1)	l	ün(1)	t
ai	b	ong(2)	l	un(2)	t
ao	c	in	m	u	u
an	d	en	n	üe(1)	v
e	e	uo	o	ui(2)	v
ang	f	ing	p	iou	w
eng	g	ia(1)	q	iang(1)	x
ou	h	ua(2)	q	uang(2)	x
i	i	ian	r	ü(1)	y
ie	j	üan(1)	s	uai(2)	y
ei	k	uan(2)	s	iao	z

其中“一码两韵”的七个字母，有下角注（1）的为第一韵，下角注（2）为第二韵。第一韵音只用在j、q、x、n、l共5个声母之后，用在除上述5个以外的声母之后都是第二韵音。

在汉字读音中有30多个单韵母音节，为了表示这些只有韵母没有声母的音节，该方案规定用字母“o”放在声母的位置上，称为“零声母”。零声母的作用是指明后续字母是韵母音。例如代码“ob”，读为ai（“爱”音）。但是，当零声母与7个“一码两韵”的字母相拼时，规定这些字母都是发第二

韵音。同时，该方案规定了用字母“e”作为这7个字母发第一韵音时的零声母。例如代码“ol”应读作“ong”（轰的韵母）；代码“el”读作“iong”（“雍”音）；又如代码“ox”应读作“uang”（“汪”音），代码“ex”读作“iang”（“央”音）。

《通信拼音字》方案的关键之一，是在熟悉以上代码规则基础上如何正确定字。所谓“定字”就是在双拼音节之后确定第三个字母（即定字字母），以区分同音不同义的汉字。这定字字母又称为调义字母，要求在“字音”的基础上，能够准确地表示“字义”，同时尽可能兼表“声调”。

该方案的具体定字规则是，首先把24个字母（不包括r，u）分成4组，每组6个字母。每一组分别表示一个声调，如下所示：

声调	字母
第1声	a b d g j z
第2声	e p t k q c
第3声	i m n h x s
第4声	o f l v w y

在定字时，先确定待定汉字是第几声，就从相应的字母组中选取一个字母，使这个字母与待定汉字有意义联系的单字的声母相同，这就是这个待定汉字的“调义定字字母”。现举例说明：

汉字“攻”的读音是第一声，与它在意义上有联系的单字是“击”，即口语上常讲的“攻击的攻”。于是就在第一组字母中选出“击”字的声母“j”，作为“攻”字的定字字母。汉字“攻”的通拼代码就是“glj”（声母＋韵母＋调义定字字母）。

如果代码组的第三个字母是“v”或“u”，通常表示该字是化学专业用字，有时也表示无声调特征的其它字。此外，第三字母还借用了汉语拼音“zh”、“ch”、“sh”三声母的第一个字母“z、c、s”。

该方案也设置了词汇码和简码。使用词汇码和简码时，可以省略定字字母，进一步方便使用，提高汉字输入速度。用单字母输入的高频简码字有18个，它们是（以字母次序排列）：

不（b）、次（c）、的（d）、副（f）、个（g）、和（h）

一（i）、几（j）、没（m）、我（o）、其（q）、四（s）

它（t）、无（u）、之（v）、尺（w）、息（x）、是（y）

词汇码的拼写方法，根据拼音字的特点，规定两字词一般应该连写，如“战略”（vdlv）、“解放”（jjnffo）、攻击（gljjig）等。当连写会引起混淆时用短横线连接，如“这次”（ve－c）、“战斗”（vd－dhv）、“一旦”（i－ddw）等。对于三字词、四字词来说，可分解为两字词处理，如“共产党”（glwwd dfn），“解放军”（jjnffo jt），“现代化”（xd-dblhq），“共产主义”（glwwd vuei）等。

13．双拼双音字词编码法

《双拼双音字词编码法》简称《双音编码》，是北京刘卫民等设计的汉字音码输入方案。它吸取了现有几种音码方案的优点，在汉语拼音和声韵双拼规则的基础上，用人们日常习惯的常用字定字方法实现汉字编码。实践证明，对于具有汉语拼音基础的人，这是一种高效的汉字输入方法。

《双音编码》的基本方法是：声母韵母双拼定音和双音选字构词。它以国标基本集（GB2312-80）作为编码汉字集，用26个拉丁字母进行编码，每个汉字一般用4个字母表示。其代码格式是：

声母＋韵母＋词汇（或字形）双音字母

该方案也规定了声韵母的替代关系，用一个字母表示汉语拼音的声母和韵母，其中有些字母也是“一码两韵”。韵母的替代关系如下表所示：

汉拼	双音码	汉拼	双音码	汉拼	双音码
a	A	ian	T	ua	J*
ai	S	iang	N*	uai	Y*
an	F	iao	Z	uan	D*
ang	C	ie	X	uang	N*
ao	B	in	M	ui	H*
e	E	ing	Q	un	K*
ei	W	iong	G*	uo	O*
en	V	iu	L	v	Y*
eng	P	o	O*	van	D*
er	R*	ong	G*	ve	H*
i	I	ou	R*	vn	K*
ia	J*	u	U		

* 表示“一码两韵”

v 表示韵母ü

声母的替代关系比较简单，如下所示：

汉拼	ch	sh	zh
双音码	I	U	V

零声母为：A、E或O。其作用与《通拼》方案所述大体相同，兹不赘述。

《双音编码》的编码规则，是在声韵双拼定音的基础上，对常用字使用双音定字法，对非常用字则用屏幕提示法，进行二次选择。所谓“双音定字”，是人们日常生活中习惯的定字方法，有词汇双音法和字形双音法两种。词汇双音法是“以词定字”的方法，犹如口语里“中国的中”、“声音的声”等等；字形双音法是以形定字的方法，如“立早章”，“弓长张”等等。该方案的双音定字实例中，词汇双音法占绝大部分。因为用词汇双音法确定一个汉字以后，也为该词汇的第二个汉字输入做好了准备。因此可实现单字和词汇融为一体的汉字输入。现举例说明如下：

例如“状”字，声母zh，代码为V；韵母uang，代码为N。因此“状”（Zhuang）用“VN”定音，然后用词汇双音定字。“状”字联系的词汇是“状态”，“态”音为Tai，代码“TS”。于是，就得到“状”字的双音码为“VNTS”，这也就是“状态”一词的词汇码，如果正需输入该词，打一下“间隔”键即可。否则，继续输入下一字的代码，此时单字“状”即已输入机器。

对于经常作为词汇中第二个字的单字，该方案规定了双音词汇颠倒定字的新方法。例如“件”字，往往在常用词的第二个字位置上，如“器件”、“元件”、“零件”、“条件”等等。用频度较高的“器件”一词颠倒定字，就是“件器”。“件”音为“JT”，“器”音为“QI”，于是“件”字的双音码为“JTQI”。

对于不常组词的单字，规定了以下几种特殊的定字方法：

（1）以字形结构双音定字。例如“陈”字，以“耳”“东”的结构定字，“耳”音为“ER”，“东”音为“DG”，故“陈”字的双音编码为“ERDG”。

（2）只用做姓氏的单字，无组词和结构特点的，用音加“姓”定字。例如“卢”字，“卢”音为LU，“姓”音为XQ，故“卢”字的双音编码为“LUXQ”。

（3）金属名，又不能组词的单字，加“金”音定字。例如“镭”字，其音为“LW”，“金”音为“JM”，故“镭”字的双音编码为“LWJM”。

（4）对于非常用的生僻疑难字，可用双拼定音，重码提示后二次选择输入。不知其读音者查国标区位码输入。

该方案规定了“单码”“双码”作为常用字的简码。单码是用26个字母各对应一个高频汉字，尽量使其声母与该字母一致，它们是（以字母顺序排列）：

安（A）、不（B）、从（C）、的（D）、而（E）、分（F）、个（G）、和（H）、产（I）、级（J）、可（K）、了（L）、们（M）、年（N）、欧（O）、批（P）、起（Q）、人（R）、所（S）、他（T）、是（U）、这（V）、我（W）、学（X）、一（Y）、在（Z）

双码是对应国标基本集（GB2312-80）中每个音节的高频度汉字，共有398个。因数量较多不易记忆，故不一一列举。

该方案的声韵母代码键位的安排（参见图3.11），考虑到使各键声韵自拼的总数在码表中的比例最高，使得操作时同一键位两次连续击键的机会最多，这对于非专职的操作人员提高输入速度是有利的。

14. 人大码汉字词语输入法

《人大码》汉字词语输入法是中国人民大学教授陶沙等在进行了“一千万字现代汉语词语频度统计”工程的基础上，从中文

信息处理的角度出发研究设计的汉字拼音编码输入方案。该方案着力在词语的构造和使用频度上下功夫，对多音汉字给予专门处理，把字、词混合编码。形成了以词语输入为主导、以单字输入为基础、短语输入作为补充的字、词、语三级结构输入体系。稍懂点汉语拼音就能很容易地熟悉它的使用方法，操作熟练后的汉字输入速度可达每分钟300字左右。

该方案以国标基本集GB2312-80作为编码汉字集，在汉语拼音方案及声韵双拼定字的基础上，把字和词混合编码。也就是说，在一种输入状态下击键4次，选中的有可能是词或短语，也有可能是字。使用者掌握了输入方法后就很容易区分使用，而不需要转换输入状态（如“单字”、“词语”状态），有利于提高输入速度。

《人大码》方案采用标准键盘的26个拉丁字母输入汉字，并按照语音原理和字母使用频率确定语音信息的压缩形式和键位排列规则，键盘语音键位图参见图3.10（在3.4节中）。键位图采用声韵双拼体制，每个音节一律用两个字母表示，前一个字母表“声”，后一个字母表“韵”。凡一个音节恰好是两个字母就用原形，例如ba（吧）、en（恩）等。多于两个字母的则用替代法加以压缩，例如gai（该）压缩为gs，shuang（双）压缩为in。《人大码》的声韵母替代关系如下表所示：

汉语拼音	人大码	汉语拼音	人大码	汉语拼音	人大码	汉语拼音	人大码
a	A	ua	Q	O	O	ie	Z
ai	S	ia	Q	uo	O	iao	X
ei	D	uai	W	ong	P	iu	C
ao	F	üe	W	iong	P	ian	V
ou	G	e	E				
ng	G	ui	R			in	B

汉语拼音	人大码	汉语拼音	人大码	汉语拼音	人大码	汉语拼音	人大码
an	H	ü	R			iang	N
		uan	T			uang	N
en	J	üan	T	ZH	V	ing	M
ang	K	un	Y	CH	U		
		ün	Y				
eng	L	u	U	SH	I		
		i	I				

声母替代的只有表中3个（ZH、CH、SH），其余都与汉语拼音的声母一样。

声调用阿拉伯数字表示：1、2、3、4、0分别表示阳平、阴平、上声、去声、轻声。

对于只有韵母的零声母音节，两个字母的仍用原形；多于两个字母的用替代关系压缩为两个字母；只有一个字母的则重复一次，补足两个字母。例如：ang（昂）→AG；ou（欧）→OU；a（啊）→AA等。

《人大码》的输入方法分为单字输入和词语输入两种。

·单字输入方法·

《人大码》方案将国标基本集中的6763个汉字分为三大类，即独用字、不独用字、不认识的字。

（1）独用字，是指既能构词又能单独使用的字，如“大、小、人、手、水、电、走、跑、说、高、低、很、不、了、吗”等，常用的约1700个字。按照三种情况输入：

①击一键——以“声母”选字的有26个，它们是：

爱(a)、不(b)、从(c)、的(d)、而(e)、分(f)、个(g)、和(h)、是(i)、就(j)、可(k)、了(l)、

每(m)、你(n)、欧(o)、派(p)、起(q)、人(r)、所(s)、他(t)、出(u)、着(v)、我(w)、向(x)、一(y)、在(z)。

②击两键——以“声、韵”选字的有412个，例如：

按(an)、被(bd)、才(cs)、得(de)、二(er)、发(fa)、

各(ge)、很(hj)、上(ik)、开(ks)、来(ls)、同(tp)。

③击三键——以“声、韵、调”选字的有约1200个，例如：安(an1)、寸(cy4)、儿(er2)、

共(gp4)、使(ii3)、考(kf3)、

美(md3)、评(pm2)、入(ru4)、

停(tm2)、位(wd4)、药(yf4)等等。

（2）不独用字，是指只能构词，除作人名、地名外一般不单独使用的字，如“绩、卑、伟、习、荣、拉、蚯、葡”等，常见的约2000个。输入时，击四键：“声韵调V”，例如：

毕(bi4v)、希(xi1v)、督(du1v)

芳(fk1v)、培(pd2v)、衡(hl2v)

宁(nm2v)、蔼(ai3v)、启(qi3v)

坦(th3v)、控(kp4v)、域(yu4v)

等等，遇同音字，用屏幕提示进行二次选择输入。

（3）不认识的字，约3000个，但使用频度只占0.01%。用笔形代码：横(h)、竖(i)、撇(p)、点(d)、折(v)，按笔顺依次取四码，取每个字的第1、2、3、末四笔。不足4码时补0。

例如：乜（折、折）→vvoo

毛（撇、横、折）→phvo

刈（撇、点、竖、竖）→pdii

犄（撇、折、撇、横）→pvph

需说明两点：

①“提”笔归入“横”，“竖钩”归入“竖”，“长撇”和“短撇”都归入“撇”，“捺”笔归入“点”。

②一个字是否认识，因人而异，这里有一个几百字的“模糊区”。为此该方案备有兼容码，按音或形都可输入，例如：“祠”字，按音为ci2v；按形为dvih，两者均可。

·词语输入法·

《人大码》方案的词语输入法分为双字词语、三字词语、四字词语和五字以上词语四种方法。

（1）双字词语输入法

进行双字词语输入时，取每个汉字字音的声母和韵母，击“声、韵、声、韵”四键。例如：

爱好(aihf)、办法(bhfa)、保证(bfvl)

产品(uhpb)、可以(keyi)、每年(mdnv)

希望(xiwk)、然而(rher)、这一(veyi)

（2）三字词语输入法

进行三字词输入时，取前两个汉字字音的声母和第3个汉字字音的声韵母，击“声、声、声、韵”四键。例如：

按规定(agdm)、比如说(brio)、不能不(bnbu)、

电视台(dits)、负责任(fzrj)、国内外(gnws)、

会不会(hbhr)、进一步(jybu)、看起来(kqls)、

另一个(lyge)、取决于(qjyu)、为什么(wime)。

（3）四字词语输入法

进行四字词语输入时，取每个汉字字音的声母，击“声、声、声、声”四键。例如：

爱国主义(agvy)、本职工作(bvgz)

对外开放(dwkf)、既然如此(jrrc)

换句话说(hjhi)、实事求是(iiqi)

如上所述(risi)、 切实可行(qikx)

（4）五字以上的词语输入法

进行五字以上词语输入时，取第“1、2、3、末”四个汉字的声母，击“声、声、声、声”四键。例如：

中华人民共和国(vhrg)

据不完全统计(jbwj)

会见时在座的有(hjiy)

在进行词语输入时，遇到有重码现象，包括二、三、四和五字以上词语混合重码的情况时，可用屏幕提示进行二次选择输入。

以上介绍了14种常用的汉字输入编码方案，是以基本笔画编码(第1～3种)、部件编码(第4～9种)、音形码方案（第10～11种）和音码方案（第12～14种）的次序排列的。由于各种条件所限，还有许多好方案没有收入，一些优秀的方案还会在实践中不断涌现出来，有待今后继续增补介绍。

如前所述，各种编码方案都各有长处和不足。现有的几种比较优秀的方案也还需要在实践中不断地优化提高，以满足日益广泛的汉字信息处理、通信和办公自动化事业发展的需要。

3·4 汉字输入键盘

如前所述，目前汉字终端最基本的汉字输入手段是键盘输入。我们在上两节中已讨论了汉字键盘输入的“软件”——汉字编码技术，现再进一步讨论它的“硬件”，即汉字输入键盘。常用的汉字输入键盘有标准键盘、整字键盘和专用键盘三

种。

1．标准键盘（小键盘）

标准键盘的全称是标准字母数字键盘。在汉字终端中，与其它两种键盘相比，标准键盘的键位少、体积小，故通称为小键盘。它实质上是一种有几十个字符键位组成的开关矩阵装置。通过按键操作，可以把各种字符（字母、数字、符号）信息变成所在键位的 x 、 y 座标值，然后转换为相应的电信号（字符代码）输入机器。

由于标准键盘的结构简单、操作方便、价格便宜，已经广泛应用于各种计算机和通信系统的终端设备上，作为实现人机对话和数据输入最简便的工具。在汉字微机和汉字通信领域，目前应用最多的汉字输入手段也是通过标准键盘的汉字编码输入。因为应用这种已经大量普及的标准键盘，可以不必另行购置其他专用设备，可靠性好，投资省，有利于汉字终端的普及和推广。

（1）标准键盘的盘面键位布局

标准键盘的盘面键位布局，是指各种字母、数字、符号键的位置排列。这是一个直接关系到键盘操作效率高低的重要因素。

众所周知，键盘的基本操作是用手指按（击）键。为了提高操作效率，要求用双手操作，并使每个手指有明确分工。根据食指、中指、无名指和小指的灵活程度以及打上下排键时的手臂活动难易，划分键位区，使双手十指“各得其所”，充分发挥其效能，这是标准键盘操作中的指法训练内容。对于键盘的键位排列设计来讲，怎样安排才能使操作者感到最方便，这

是汉字输入编码方案设计的重要内容。由于目前各类编码方案的键位排列，都是在标准键盘的基础上实现的，因此有必要首先讨论一下标准键盘的盘面键位布局，然后分别讨论各类编码方案的盘面键位布局问题。

标准键盘的盘面布局，大致可分为基本型和扩充型两类。

基本型键盘的盘面布局如图3·8所示。图中(a)是斜键盘布局，(b)是正键盘布局，两者的键位排列完全相同，只在越排操作时手臂活动范围略有差别。

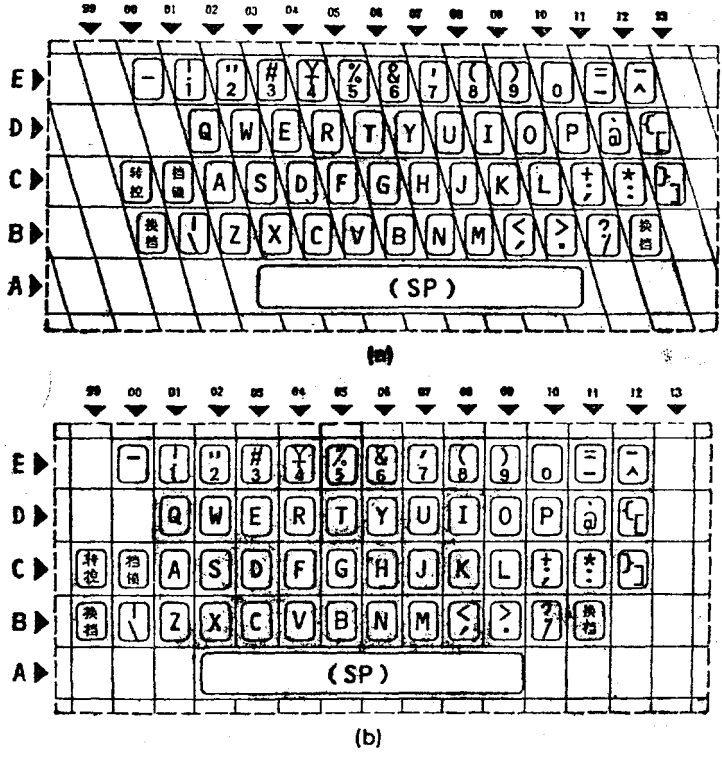


图 3·8 标准键盘的盘面布局（基本型）

基本型键盘盘面，通常设置50多个键位，分别排列成5行。除了第一行（图中的A行）只有一个间隔键（SP）外，其余各行（B行～E行）一般安排10多个键位，可以分为三个大区，即字母键区、数字键区和符号功能键区。

字母键区安排了26个拉丁字母，分成三行，即从B行（第2行）到D行（第4行）的01列到10列之间，是标准键盘盘面的主体区域。这26个字母键的键位排列次序，是1873年创制英文打字机时根据字母使用频度和手指合理负担原则确定下来的。尽管后来发现其排列不太合理，但已经沿用一百多年，“约定俗成”，很难改动了。

数字键区安排0～9十位阿拉伯数字键，置于键盘盘面最上边的E行（第5行），以从左到右数字递增次序排列，但“0”键在最右边。

符号及功能键区安排在盘面的两侧以及数字键的上档。通常设有10多个常用符号及必需的操作功能键，如“回车”“换行”“换档”（大小写）等等。随着机器功能的扩大和完善，键盘功能键的设置常随机器型号不同而各不相同，图中所列的仅是其中的一种。

间隔键(SP)又称为空格键，是专为输入字符之间的间隔（空格）而设置的，安排在最下边A行的中间，单独占据一行的位置，通常用双手的大姆指击键操作。在汉字输入编码中，间隔键常用作结束一个汉字代码组的标志符。

基本型键盘的键位数较少，适用于汉字电传机、简易型汉字终端的汉字及字符输入。

随着计算机和终端设备应用范围的扩大，需要增加更多的符号及功能，如文字编辑中的增字、减字、增行、减行等等，原有的基本型键盘已经不能满足要求，于是产生了各种扩充型

键盘。一种典型的扩充型键盘盘面布局如图 3·9 所示。

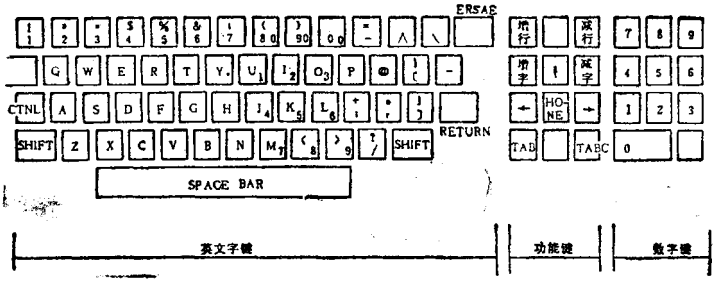


图 3·9 扩充型标准键盘盘面布局

扩充型标准键盘是在基本型键盘的基础上，又增加了一个功能键区域和一个数字键区域。基本型键盘原有的字母键区和数字键区位置排列不变，以保证熟练操作人员的快速输入。新设的功能键区，主要是供文字编辑和制表格时的“增、删”、“查找”、“存储、读出”、“表格设置”等功能及参数的输入之用。新设的数字键区是为了适应日益增多的数据输入的需要。把10位阿拉伯数字集中设置，这样可以用一手(右手)输入数据，腾出另一手翻阅单据或资料，以提高输入效率。当然，对于只需10位阿拉伯数字的汉字基本笔画输入法（五笔画、八笔画）使用集中设置的数字键输入，也是比较方便的。

为了使汉字输入也与西文一样用双手按键操作输入，甚至在速度上超过西文输入，仅仅用10位数字键是远远不够的。因此，必须进一步讨论标准键盘用于汉字输入时的盘面布局，使之与汉字编码方案相适。

用于汉字输入时的标准键盘盘面布局问题，大致可分为音码输入方案和形码输入方案两部分。

① 音码输入时的盘面布局

我国的汉语拼音方案规定直接采用国际通用的26个拉丁字母做为汉语拼音字母。因此，当使用音码方案输入汉字时，不论是汉语拼音输入还是双拼输入，都是采用标准键盘字母键区的26个字母，不必增加新的字母或图形符号。但是这些字母键的排列位置却不同。各种音码方案的设计者，从不同的角度出发，已经推出了适合汉字输入特点的几十种双拼编码键盘盘面布局方案。方案虽多，归纳起来大致可分成两大类型：一类是以标准键盘盘面布局为基础的兼容性方案，另一类是从汉语语音特点出发的合理化方案。

第一类方案考虑到标准键盘的基本布局已经通用多年、已被许多人熟悉掌握的实际情况，规定在音码双拼输入方案中，凡是与原有拉丁字母相同的声母，都按照标准键盘的排列位置不变。对那些与原有拉丁字母不同的声母和韵母，或者以汉语拼音方案中不用的字母(如字母“V”)代替，或者按照“一键二义”的原则由已用的字母“兼任”，使声、韵合键。例如在Q键上增加复韵母ing，在W键上增加复韵母ei等。这种一键二义式键盘，操作时不需经过“换档”，机器即能自动识别，使用比较方便。

这类方案的最大特点是兼容性好。由于大多数声母键键位排列与标准键盘的字母键键位排列相同，使熟悉英文打字的操作人员容易掌握，并能达到较高的汉字输入速度。因此，在目前众多的音码双拼输入方案中，大多数采用这类声韵合键的盘面布局方案。

这类方案的韵母键位排列则各不相同，这是因为各种编码法的韵母替代关系规定各异，还没有统一标准的缘故。现举两种音码输入方案的键位盘面布局为例（见图3·10和3·11），供读者比较参考。

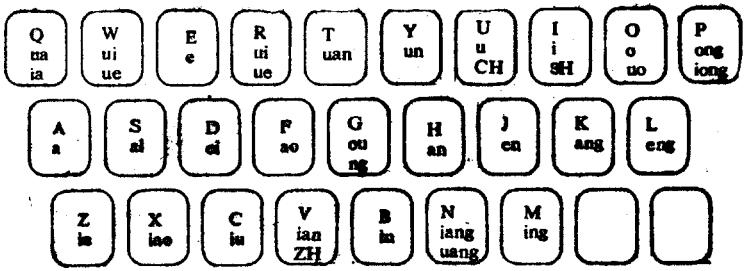


图 3·10 以标准键盘为基础的双拼键盘盘面布局之一
（人大码键位图）

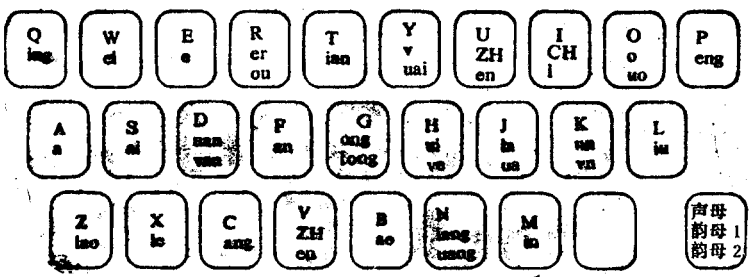


图 3·11 以标准键盘为基础的双拼键盘盘面布局之二
（双音编码键位图）

图3·11中，双音编码的韵母键键位设计时，考虑到非专业操作人员不习惯双手轮流击键操作的特点，尽量使一个汉字音节在双拼时连击同一键位的概率最大。

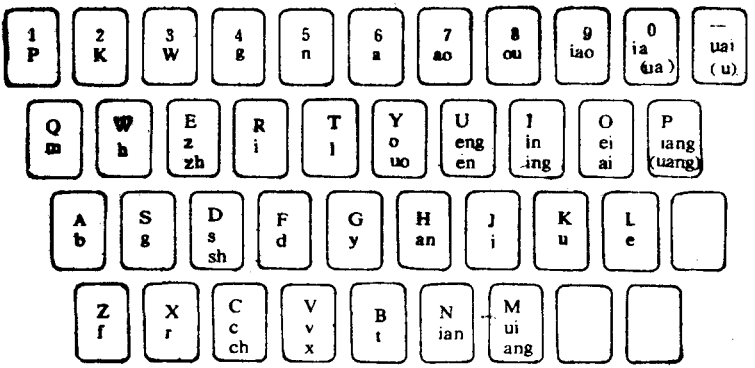
另一类方案是从汉语拼音方案的字母使用频度与英文字母使用频度不同的特点出发，按照双手操作键盘时手指负担合理化的原则，对标准键盘的键位排列进行重新设计的一种声韵分键键盘布局，又称为合理化方案。

总结外文键盘打字的经验，双手轮流击键操作时节奏感强，长时间操作不易感到疲劳。而一个汉字的双拼音节总是一声一韵。因此，应把声、韵母键的键位分区布局，声母键安排在盘面的左边，韵母键安排在盘面的右边。这样的键盘布局，能使汉字双拼输入时双手轮流击键，以达到减轻操作疲劳、提高输入速度的目的。而且，把声韵母键分区安排，还可以把读音相近的声母或韵母安排在同一个键位上，如果后随着区分力强的信息时，可以不必用“转档”键操作。这样一来，能够做到把某些地方口音中容易发生混淆的声母或韵母，例如“Z”与“ZH”、“S”与“SH”、“C”与“CH”，以及“en”与“eng”、“in”与“ing”、“n”与“l”等等，都放在同一个键位上，操作时可以当作同一字母看待，大大有利于音码输入方案在方言地区的推广应用。

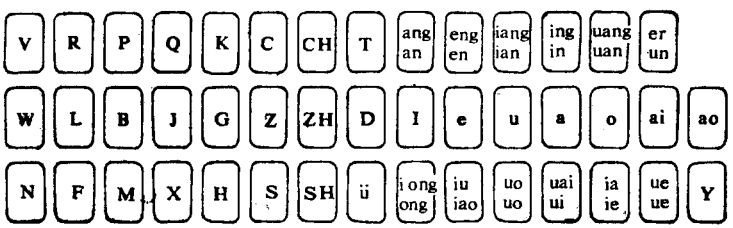
键盘布局合理化方案的另一个内容是，在声、韵母键位区内的具体键位安排时，应按照汉语拼音字母使用频度高低为序合理布局。测试试验结果表明，在进行键盘打字操作时，食指的动作最灵活且省力，中指次之，小指和无名指最差。对四排键位打键时，手停的一排键（即基本及还原位置）最易打，上一排较难些，下一排又再难些，而最上排则最难。因此，应把汉语拼音方案中使用频度最高的字母放在手停一排的中间位置，然后依使用频度向上下和两边放射排位。图3·12（a）和（b）就是这种声韵分键合理化键盘盘面布局图。类似的方案还有不少，限于篇幅，这里仅举两种供读者参考。

由图可见，这类键盘布局与标准字符键盘布局完全不同。这样虽然失去了两者的兼容性，但从长远观点来看，还是很有实用意义的。

② 形码输入时的盘面布局



(a) 四排键布局



(b) 三排键布局

图 3·12 两种声韵分区双拼键盘盘面布局

如前所述，形码输入时是把字形拆分成若干个部件或笔画，再按一定规则顺序击键输入的。因此，形码输入时的键盘盘面标记和布局与标准键盘盘面不同，而且随着不同的编码方案而异。几乎有一种编码方案，就有一种盘面布局方案。现选择几种具有代表性的形码输入键盘盘面布局介绍如下：

一类是基本笔画编码输入法的键盘布局。

基本笔画编码输入法是以“笔画”作为构成汉字字形的

“字母”，常用的有“五笔画”和“八笔画”两种。因此，必要时只需10个数字键的小型键盘便能实现汉字输入。虽然这样的键盘输入速度相当慢，但可以减少设备的体积和成本，在某些需要小型化设备的场合还是有用的。在大多数使用场合下，基本笔画编码法仍然用标准键盘来输入汉字。这样可以把一些常用的笔画组合安排在各个空闲的字母键或数字键键位上，以减少输入时的按键次数，并能进行双手操作，使输入速率大大提高一步。

图3·13是《基本笔画笔顺编码法》的键盘盘面排列。它把五种基本笔画安排在字母键区中排D、F、J、K、L五个键位上。把其余21个字母键和10个数字键分别安排了23个常用的笔画组合（部件）键和8个高频字简码键。键盘键位数的选择是基于这样的原理：键位数少，输入一个汉字的平均按键次数

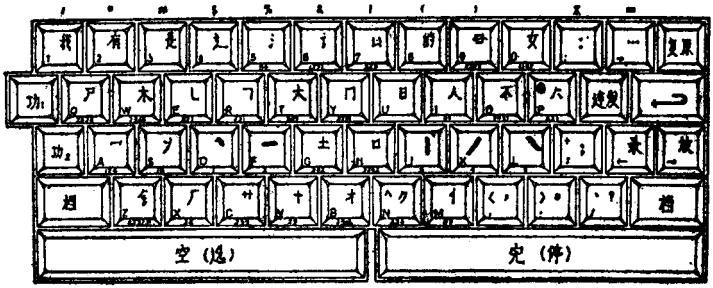


图 3·13　基本笔画笔顺编码法36个基本键和多画键的键盘布局

多；键位数增加，平均按键次数减少。按该方案的编码规则输入一个常用的“哲”字，如果只用五个基本笔画键输入，需要打12次键。把键位数增加到30个，平均按键次数可以减少一半，“哲”字只需打键6次便可输入。同时，使标准键盘上的键

位得到充分利用。

其它几种基本笔画编码输入法，如五笔画和笔形编码法（八笔画）都以类似方法安排盘面布局。

另一类形码键盘布局是部件编码法键盘布局。

如前所述，部件编码法是以部件或字元来作为构成汉字字形的"字母"。但是，汉字的构字部件有几百个，即使归并整理后还有一百多个。要把它们安排到标准键盘的几十个键位上，就得根据选定的键位数把部件分类合并，才能使它们"各得其所"。

部件编码键盘的键位数大体有26键型、36键型和45键型三种。26键型方案是把全部构字部件归并到26个字母键键位上，这种方案的好处是在进行汉字输入操作时，双手手臂的活动范围较小，有利于减少疲劳，提高打键输入速度；缺点是每个键位上安排的部件数较多，不容易把它们都标示在键帽上，就要增加操作时的记忆量。一种典型的26键部件编码键盘布局如图3·14所示。这是前面介绍的《五笔字型》和《五笔桥》编码方案在88年5月制定的标准键帽键位图。

45键型的键盘布局方案，是把一百多个构字部件分配到除几个特殊功能键外的45个键位上。也就是26个字母键位、10个数字键位和9个符号键位都用来作为汉字部件的输入键，这样，每个键位上只需安排不超过4个构字部件。一种典型的45键汉字部件编码键盘布局是图3·5所示的"《联合45-3拼形输入法》部件键位图"。这种安排的好处是操作者对部件的键位安排明了直观，易于操作；而且减少了学习时记背笔画部件的负担。其缺点是操作键盘时手臂的活动范围增大，时间一长易感疲劳；而且，在熟练后要进一步提高汉字输入速度会受到一定限制。

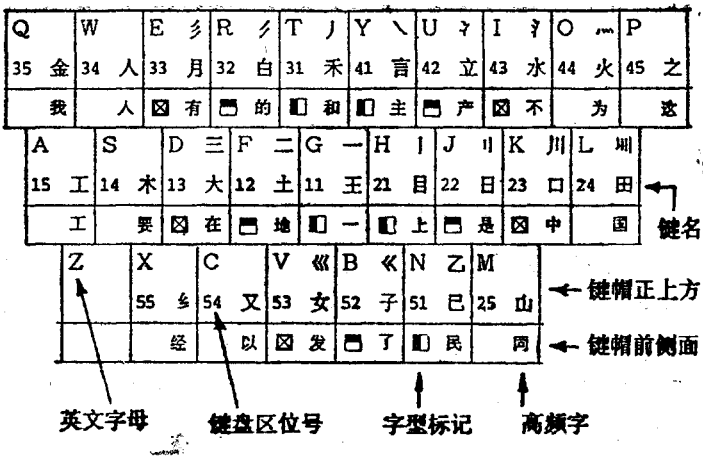


图 3·14 五笔字型标准键帽键位图

36键型的键盘布局方案是把构字部件分配到只限于26个字母键和10个数字键键位上。湖南欧阳松设计的《CK码汉字输入方案》就是采用36键型的键盘布局。

以上讨论的形码键盘盘面布局，不论是何种类型，有一个共同的特点，就是看上去一大片各种笔画和部件，不免使人眼花缭乱，不知从何下手。其实，任何部件编码键盘布局都是经过设计者反复推敲、周密考虑安排的；有些方案还使用计算机辅助优化设计，是有规律可循的。

首先，这些盘面布局大多是以部件的起笔笔画分类、分区安排的。一般是以“横、竖、撇、点、折”分类，将整个盘面分成5个区。将具有相同起笔笔形的部件安排在同一区内，以方便操作者的记忆和查找。

其次，对同一区内键位按照部件的实用构字频度高低合理分配。通常把实用构字频度高的部件安排在中排的中间几个

键位上，由双手食指操作。然后按照由中间向两边，先中排、再上下排、最后最上排的次序进行键位分配，力求使汉字输入操作方便、快速、省力。

最后，要根据部件的字形结构不相容性，恰当地把几个部件安排在同一个键位上。这是实现部件编码键盘盘面正确布局的关键。所谓部件的字形结构不相容性，是指这百多个部件中，并不是任意部件都可以相互搭配构成字形的。而且许多部件的构字位置也不是随意的，有些部件只能出现在字形的左边，有些部件则只能出现在字形的右边。于是人们可以从部件构字的特点中找到这样的规律：当操作者按下一个键时，使机器自动识别您想要输入这个键位上的哪个部件，不必附加“换档”或“选择”之类的操作。

以《HPX拼形输入法》为例，如输入“结”字，只需按照编码规则按“LRKH”4键（“H”是结束键）。我们可以看到在“L、R、K”3个键位上，各有4个部件，但是如果选取其它部件，都不可能按照输入的编码顺序构成字形，因而只能选取“纟、士、口”3个部件构成“结”字。其它部件编码盘面布局也都具有这种能力，因而大大方便了部件编码法的汉字输入。

读者也许要问：“如果只按一次键，机器能知道要输入键位上的哪个部件吗？”当然是不可能的。所以编码方案都规定了单独打一键时输入键面上某个部件或单字时的操作规则，这在上节讨论各种编码方案时已经举例介绍，不再赘述。

（2）标准键盘的按键结构及其特点

标准键盘的机械结构比较简单，盘面上是排列成矩阵的几十个键，每个键的下部实际上是一个开关装置。通过按键动

作，开关闭合一次，机器便将这个键位所在矩阵的座标量读出，然后转换成相应的字符代码信号。

按键开关的种类很多，可分成机械触点式和无触点式两大类。

机械触点式按键的结构简单，体积小，便于大规模生产，价格低廉；在使用中耗电量比较省，对环境温度变化的适应性强；可靠性好，能正常动作10^7次以上，不需经常维护。但是机械触点制造工艺复杂，对材料的要求比较高。

机械触点式按键有导电橡皮式、微动开关式、笛簧管式几种。

导电橡皮式按键是一种应用广泛的机械触点式按键，其结构如图3·15所示。

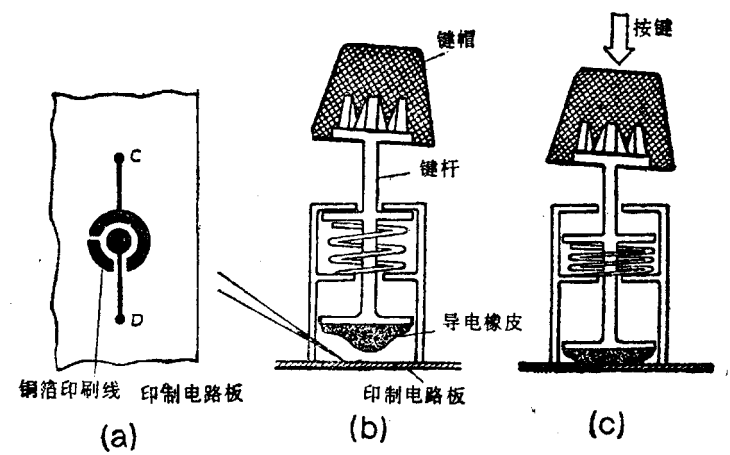
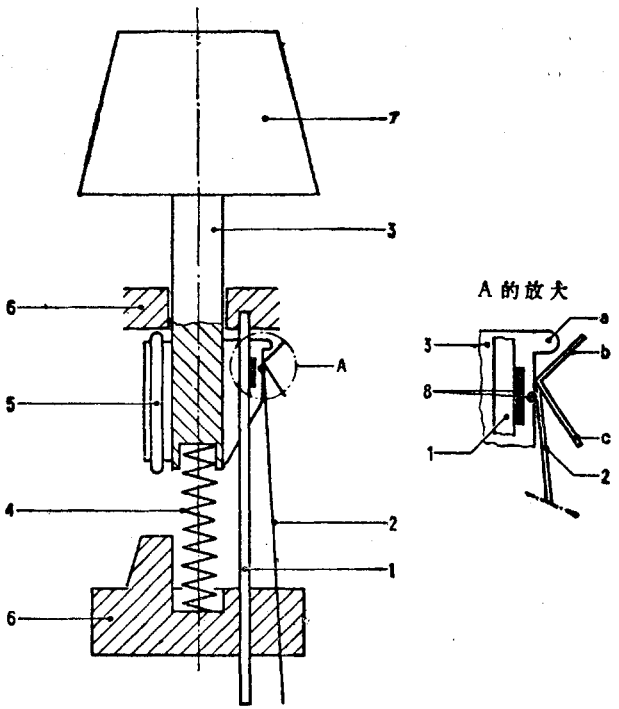


图 3·15 导电橡皮式按键结构图

由图可见，在键杆的上端是标有字符的键帽，下端是可以上下活动的导电橡皮。当按下键帽时，导电橡皮顶部同时压住印刷电路板上的两个金属导体C、D，使它们通过导电橡皮连

接起来，达到按键开关闭合的目的。导电橡皮的电阻值，在键内配合的直径两端测量应小于200Ω。

微动开关式按键是利用手指击键时的力量，使金属接点接触，从而达到使按键开关闭合的目的。微动开关式按键的种类有多种，一种典型的按键结构如图3·16所示。



1——触点板，　2——触点弹簧，3——键杆，
4——复位弹簧，　5——阻尼环，　6——外壳，
7——键帽，　8——触点，　a——键杆凸点，
b、c——触点弹簧限位片

图 3·16　微动开关式按键结构图

由图可见，键杆 3 由于复位弹簧 4 的张力作用，平时保持在上面的位置，因而使触点板 1 与触点弹簧 2 断开，按键开关不通。当按下键帽时，附于键杆一侧的凸点a推动触点弹簧使触点 8 闭合，开关接通。

笛簧管式按键是一种利用磁力驱动机械簧片接触的开关装置，其结构如图3·17所示。

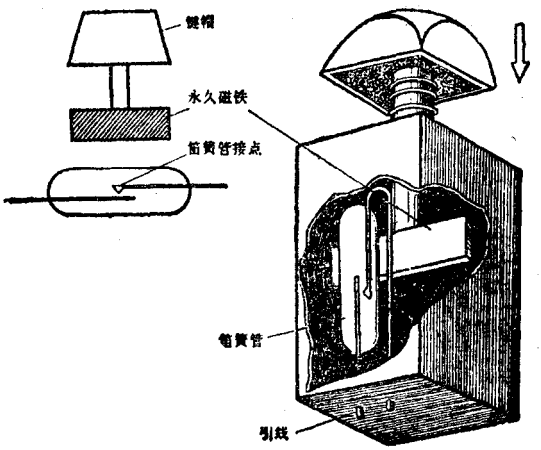


图 3·17　笛簧管式按键结构图

图中笛簧管是一个充有惰性气体的玻璃管，管中封装着两片带有触点的弹性黑色金属片。当按下键帽时，装在键杆下端的一块永久磁铁靠近笛簧管，在磁力作用下使金属片动作，达到使按键接点接通或断开的目的。由于按键接点被密封在玻璃管内，不易被污染或氧化，因此使用寿命很长，接点能正常工作达5×10^{7}～5×10^{8}次。

无触点按键是利用手指击键时键杆的上下运动促使一种物理量（如电、磁或光的参数）改变的原理来产生开关信号的装

置。由于没有直接的电气触点，这类按键的动作可靠性好，故障少，工作寿命可达10^8次以上。但这类按键的结构比较复杂，动作时耗电较多，对环境温度变化的耐受能力较差。

无触点式按键有电容耦合式、磁耦合式、霍尔效应式及光电耦合式等数种。

电容耦合式按键是利用按键时键杆向下运动而改变键底电容量的办法来产生按键信号的装置，其结构及工作原理如图3·18所示。

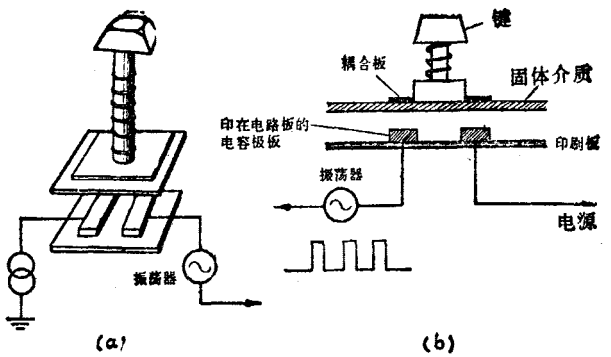


图 3·18　电容耦合式按键的结构及工作原理图

如图3·18所示，当按键时，装在键杆下端的耦合板和固体介质向下移动，压在印刷板上电容器的两个突出的极板上，使电容器的电容量增大到120pF左右。该电容器是作为外接振荡器的振荡元件，当电容量增大到额定值时，即产生振荡信号。该振荡信号使触发器形成一个电脉冲，表示该键被按下一次。这个电脉冲还将用来形成键码信号。由此可见，电容耦合式按键本身只产生一个电容量，而不直接输出电信号，因此工作寿

命比较长。

霍尔效应式按键是近年来得到迅速发展和广泛应用的一种键盘按键。它是应用霍尔效应原理产生开关信号的，其结构如图3·19所示。

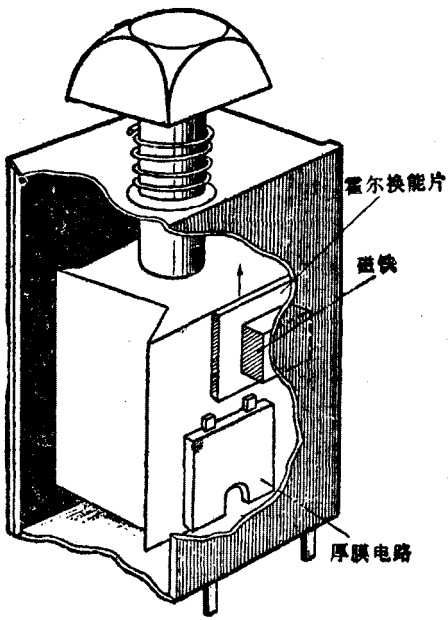


图 3·19　霍尔效应式按键结构图

所谓“霍尔效应”是E·H·霍尔在1879年发现的一种物理现象。就是当一块矩形导体或半导体材料置于磁场中时，这个材料会产生一定方向的电流和电势。图中的“霍尔换能片”就是一块半导体材料。当按动键杆时，附在键杆上的永久磁铁靠近霍尔换能片。于是该换能片输出一个电压，送入装在键壳内

的厚膜电路，经整形后便得到一个脉冲开关信号输出。这种按键的可靠性好，而且电平输出有高、低两种可供选用，使用比较方便。

下面简要讨论一下按键接点的抖动问题。

任何一个机械接点在接触过程中都会产生抖动现象，按键接点也不例外。当操作者按下一个机械触点式按键时，键底触点开关并不是立即闭合，而要经过一段时闭时开的抖动时间。图3·20是一个键接点在按键时触点电阻变化的典型抖动曲线。图中所示的电阻值R_1和R_2是接点闭合前后的稳定电阻值。当按下键时，要经过t_1至t_2的前沿抖动时间；抬起键时，要经过t_3至t_4的后沿抖动时间。一般抖动时间在10毫秒以内。按键接点的抖动，往往会使机器产生错误识别，造成输入差错。因此，必须采取相应措施加以解决。

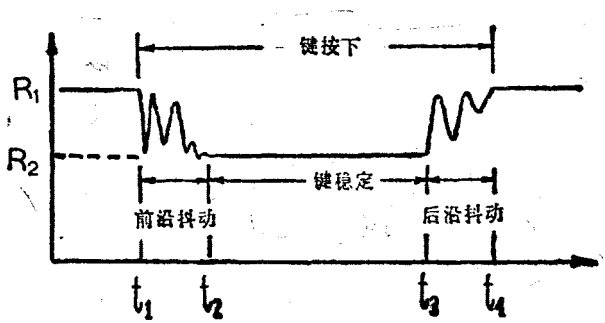


图 3·20 按键时接点电阻值的抖动曲线

克服接点抖动的方法有两种：一种是用软件延迟的方法。就是当按键以后，CPU不立即进行查询，而是经过一段延迟时间，等待按键状态稳定以后再读取该键的状态信息。延迟时间的长短根据接点抖动时间用程序设置。另一种是用硬件滤

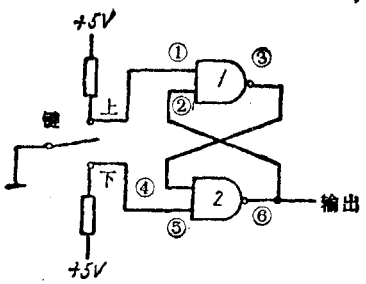


图 3·21 防止抖动的电路

波的方法。例如可用一个R-S触发器组成防抖动电路，如图3·21所示。图中不论键位置向上或向下闭合时，只要与非门1的输入端①出现一个低电平，通过输出端③使与非门2的输入端④为高电平；于是又通过输出端⑥使与非门1的另一输入端②为低电平。这就保证了在接点闭合抖动过程中，输出端⑥始终保持稳定的高电平输出。

（3）键盘编码的基本工作原理

键盘作为输入设备的一项基本功能，就是把操作者的按键动作变换成相应于该键位的电码信号（或称为键码信号）。如上所述，按键动作是接通位于键盘矩阵中某个键位的开关接点。机器在读取了这个接点的座标量之后，再将它变换成键码信号的过程，就称之为键盘编码。

标准键盘实现键盘编码大体有两种方式，一种是软件方式，就是通过程序的方法来实现编码；另一种是硬件方式，就是通过专用的编码电路来实现键盘编码。

首先讨论用软件方式实现键盘编码的简单工作过程。

用软件方式实现编码所使用的键盘称为非编码键盘，这种键盘的电路结构如图3·22所示。

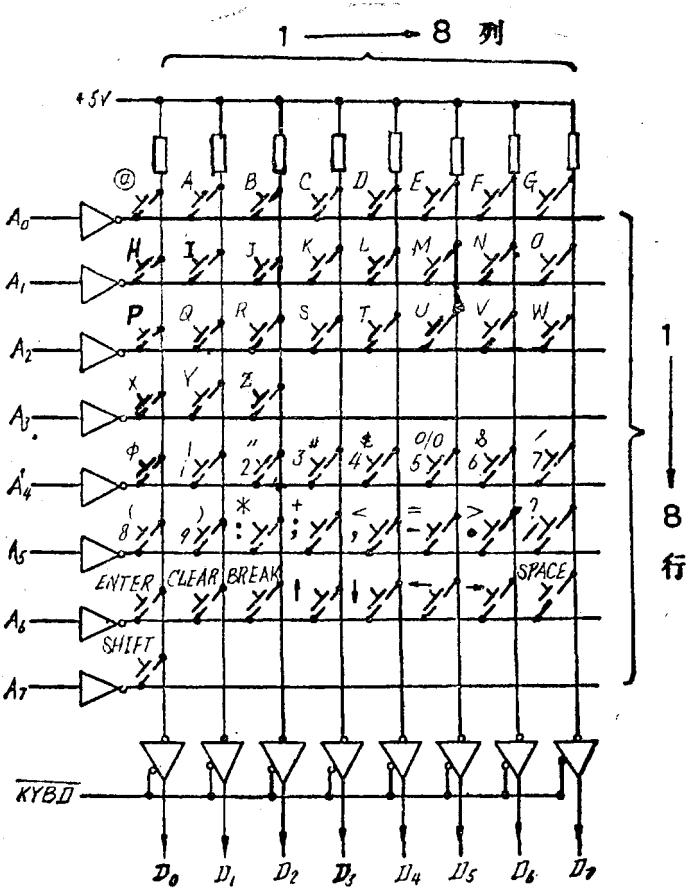


图 3·22 非编码键盘电路结构图

由图可见，这是设有53个键位的 8 × 8 开关矩阵，每个键位是一个单投常开接点，用来表示某一字母、数字、符号或操

作功能，其电路结构十分简单。各种键码信号都是通过程序查询、扫描和查表的方法产生的，其工作过程可分以下几步：

① 查询是否按键。CPU输出高电平信号到A_0～$A_7$8根地址线上，通过各自的输出反相器使各行线为低电平（接地）。然后使$\overline{KYBD}=0$，打开输入反相器，CPU读入D_0～$D_7$8根数据线上的输入信号。如果输入信号全为0，表示没有按键；反之，如某一列线的信号为1，则表示有键按下，立即转入以下几步操作。

② 确定按下的键位在哪一列。CPU通过对数据线D_0～D_7的检测，便能确定按下的键在哪一列。例如测知$D_3=1$，表示按键在第4列。

③ 确定按下的键在哪一行。CPU向地址线A_0～A_7依次输出高电平。例如先使$A_0=1$，其余各线为0，经过反相器后，第一行为0，其余各行为1。再从D_0～D_7线读入数据是否全是0。若是全部为0，则再使$A_1=1$，其余各线为0，依此类推。直到查出读入的D_0～D_7线中有一个为1时，对应地址线为1的行就是按键所在的行。例如，当$A_4=1$时，CPU读入D_0～D_7中有一个为1，则按键在第5行。

④ 把行号列号转换成键码信号。当确定键位的行、列号后，首先应判断该键位是功能键还是字符键。是功能键就转到相应的子程序处理；是字符键则可采用查表方法，把行、列号转换成相应的字符代码。例如已知第4列第5行键位是数字“3”，则把上述地址值转换成16进制数“33”（数字“3”的ASCII码）。

用软件方式实现键盘编码虽然电路结构简单，设备成本较低，但占用CPU的工作时间较多，速度较慢，现已很少采用。

用硬件方式实现键盘编码，就是主要采用数字集成电路实

现由键位座标量到键码信号的转换，这是目前应用较广的键盘编码方式。具体电路的种类比较多，这里以一种带只读存储器的扫描键盘（图3-23）为例来讨论一下实现硬件编码的工作过程。

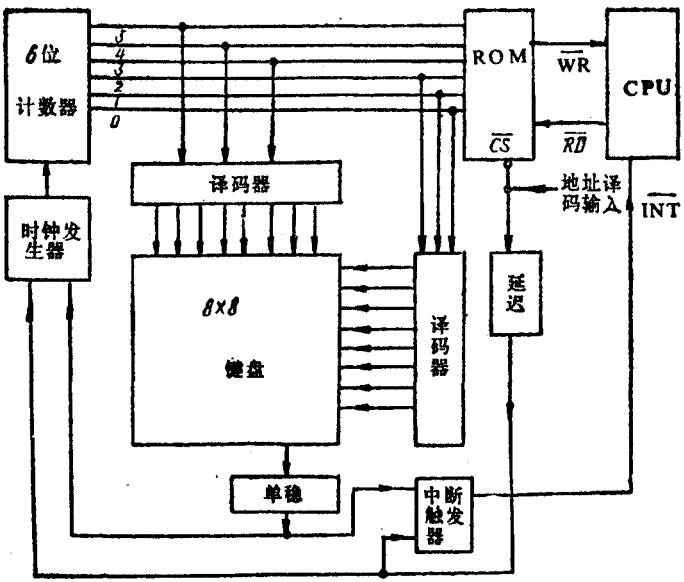


图 3·23 带只读存储器的扫描键盘工作原理框图

图中，一个 8 × 8 键盘就是具有 8 行 8 列的非编码键盘，可以提供64个键位（参看图3·22）。它通过 2 个八中取一的译码器，用 6 位计数器进行扫描。如果键盘没有键按下，扫描将随着计数器的计数循环反复进行下去。一旦有键按下，键盘立即通过单稳电路产生一个脉冲信号，使计数器停止计数和扫

描。此时，计数器的输出与按下的键位相对应：计数器的前3位对应键位的行号，后3位对应键位的列号。于是可以把计数器的输出作为只读存储器（ROM）的输入地址，控制对存储单元的读取，读出的内容就是相应键位的字符代码。

单稳电路输出脉冲的另一个作用是通过中断触发器向CPU发出中断请求信号。CPU响应中断后立即对只读存储器（ROM）发出读取指令，把字符代码读入CPU。读取指令的地址译码信号经过延迟后用来清除中断触发器和重新启动6位计数器工作，开始新的扫描过程。这种键盘的工作速度很快，当几个键几乎同时按下时，它将按照键盘扫描的先后次序依次读入。

随着LSI技术的发展，已经研制和生产出多种用于键盘扫描编码的专用大规模集成电路芯片，可供不同规格的键盘使用。例如Intel8279可编程序键盘/显示器接口芯片，可提供64键的扫描接口，还可扩展到128个键位。键盘入口设有防止按键接点抖动的电路，还设有输入信息的缓冲存储器。键盘的扫描方式、移位控制等功能均可通过编程加以改变，使用十分方便。

2. 整字键盘（大键盘）

整字键盘是以整字（即单字）设键的键盘装置，是汉字终端常用的输入设备之一。

如前所述，用标准键盘输入汉字时，由于汉字单字数量远远多于键盘的键位数，必须先进行输入编码，把汉字变成字符代码才能输入。用整字键盘输入汉字，可以免去汉字输入编码过程。这对于不熟悉汉字编码的非专职操作人员来讲，使用整字键盘输入是一种比较简便的输入手段。

整字键盘的特点是在键盘盘面上直接标出汉字，直观易学，操作简便。用它进行输入操作时，见字选字，一般人员都可以掌握。在熟悉了盘面汉字（即字表）的排列位置以后，输入速度可达每分钟40～70个汉字，而且可持续工作数小时不易感到疲劳。尤其在使用汉字字数有限、且具有某些固定操作的专用系统，如机场、车站、码头、旅馆等场合的查询服务业务中，采用笔触式整字输入方式可以大大提高终端设备的效率。

（1）整字键盘的种类和特点

整字键盘的种类很多，根据其结构方式的不同，大体可分为全键式、换位式和笔触式三大类。

全键式整字键盘的按键结构与标准键盘相似。键盘盘面是由二、三千个常用汉字及字符以一字一键方式构成的开关矩阵装置。输入操作时，先找到所需汉字或字符所在的键位，通过按键动作，使该键下面的开关接点接通，机器通过计数扫描电路检测到该键的座标后，通过查内存表或简单运算变换成该字的代码输入。

从单键操作过程来看，全键式整字键盘与标准键盘在原理上没有什么两样。但是，由于汉字的常用字数量很大，安排二、三千个盘面字，用起来往往感到不够；但这几千个按键结构组合起来，使全键式整字键盘的体积庞大、笨重，操作时按键动作比较费力，汉字输入效率很低，因此，这种整字键盘现已很少使用。

换位式整字键盘又称为主键——辅键式键盘。它是为了减小全键式整字键盘的体积、又可以适当增加盘面单字数量而设计的一种整字输入键盘。

换位式整字键盘通常设置二、三百个键位，称为主键。每

一个主键键面上标有十多个汉字。另外设置与主键键面汉字数相同数量的键位，称为换位键或辅键。用换位式整字键盘输入汉字时，先按下所需汉字的主键，再根据该字在主键上的排列位置，按下相应的换位键。例如，需要输入“们”字，先找到“们”字所在的主键键位。如图3·24所示，假定该键上共标出12个汉字，“们”字在键面的右上角位置，则应在按下主键后，再按换位键的相应键“3”，便可输入。

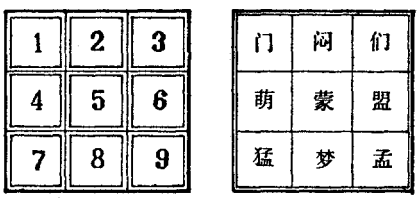


图 3·24 换位式整字键盘的键面举例

换位式整字键盘的收字数量，取决于主键键位数与每个键位键面字数的乘积。含有256个主键，每键12个汉字的整字键盘总收字数量为3072个。对于一定数量的汉字集来讲，主键键面容字数量少（这样找字时比较醒目），则要求键位数就多，键盘的体积就更增大；反之，增加主键键面字数，可以减少键位数量。但主键上的汉字过密，增加了找字的困难。而且如果收字总数不变，则键盘体积减少甚微。目前应用的换位式键盘，一般收汉字四、五千个，主键键位数为二、三百个，每个键位上安排单字 9 ～15个。

笔触式字盘，又称笔触式键盘，是近几年来迅速发展起来的一种汉字整字输入装置。

笔触式字盘的盘面是一张以一定格式印成的汉字整字表，简称字表。每个汉字占有字表上一个固定的坐标位置，字表覆

盖在一块精度要求不高的图形输入板上。当进行输入操作时，操作者用一支特殊的触笔点触字表中的一个汉字，机器通过检测触压点的坐标位置，便能将该字的代码输入。笔触式字盘的组成框图如图3·25所示。

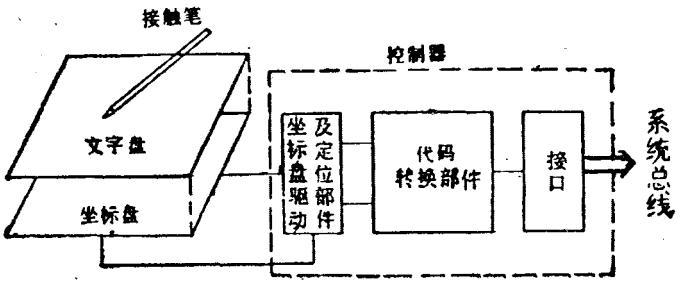


图 3·25　笔触式字盘组成框图

如图所示，笔触式字盘通常是由文字盘（即字表）、坐标盘、接触笔、控制器 4 个部分组成。控制器用来对坐标盘进行驱动扫描，当触笔触及文字盘时，它立即检测出触点的x、y坐标值，经过代码转换电路变成相应的汉字代码。然后通过接口与系统总线相连，把代码送入缓冲存储器（图中未标出）。

笔触式字盘只需一张字表及一块结构简单的坐标盘取代了数量很大的按键装置，使整字键盘的体积重量大大减少；同时用笔触代替打键操作，操作员不易疲劳，可以显著提高输入效率。因此，笔触式字盘已成为应用最多的整字输入键盘。

笔触式字盘的种类较多，按不同的工作原理区分，有静电式、电磁式、光电式、磁致伸缩式及压感式等。输入用的触笔一般分为有线式和无线式两种，有些压感式整字键盘可不用笔触，而是直接用手指触摸即可输入，也可用普通圆珠笔或圆头笔杆点字输入。

为了扩大笔触式字盘的收字数量，还设计制成了翻页式笔触键盘。就是把键盘所收汉字分成几页字表，重叠放置在坐标盘上面。每页字表可以象书页一样翻动，其翻页机构与定位计数装置相连，翻页时能自动输出页号代码。因此，只要找到所需汉字的某页字表后，仍是一次笔触输入一个汉字。翻页式笔触键盘使用多页字表，使键盘收字数量大为增加。虽然输入时需要翻页寻找，但基本免除了处理盘外字的麻烦，具有一定的实用价值。为了便于使用，翻页式键盘通常以常用字、次常用字、词组、专用术语等不同字（词）类别分页，以利查找。

综上所述，笔触式整字键盘直观、易学，操作简便。但是它有两个主要缺点：一是找字比较困难，二是与标准键盘相比，专用设备的成本较高。随着整字键盘盘面字排列的国家标准的制订和实施，找字难的问题将会逐渐减小；随着设备大批量生产后成本的逐步降低，整字键盘的应用还是大有前途的。

（2）整字键盘盘面字表的布局

汉字整字键盘的盘面字表是直接面向用户的，字表汉字及其它字符的排列布局，对键盘输入操作速度影响很大，并与键盘的结构特点密切相关。盘面字表布局问题的主要内容包括：整字键盘盘面字的分区、盘面字数的选择和排列、盘外字的处理方法等。

① 整字键盘盘面字的分区

目前国内外使用的整字键盘种类较多，盘面字的排列规律各不相同，但大致都把盘面分成以下几区：即汉字区、非汉字字符区、功能键区和保留（机动）区。

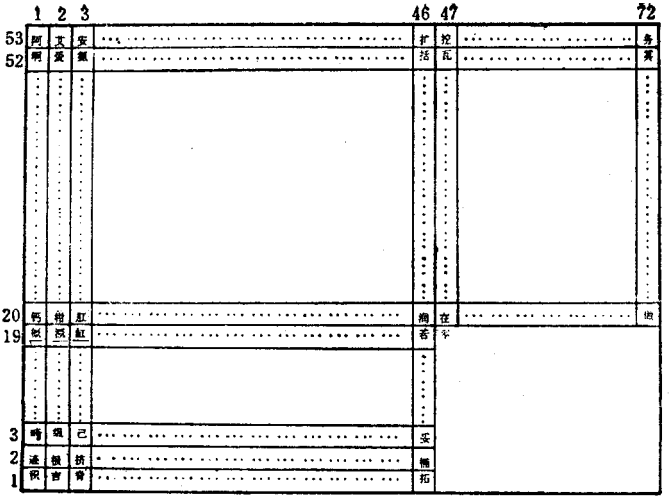
汉字区就是字表，它是整字键盘盘面的主体部分。字表的

收字容量和字的选择，应符合我国常用汉字的使用规律，并考虑到设备制造的方便。根据国标汉字基本集（GB2312-80），一级字3755个为常用字，覆盖率达99%，其中的最常用字1987个的覆盖率可达95%以上。因此就通用型的整字键盘来说，汉字区的字数大体在二千至四千个左右。这样的收字量，既保证了大多数情况下的使用需要，又不致使盘面结构过于庞大，也有利于操作时的查找。

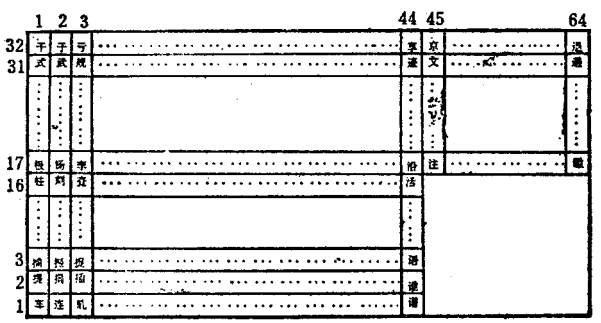
非汉字字符区、功能键区和保留区通常安排在盘面字表的右下方。非汉字字符区的内容应包括标准键盘盘面字符的全部内容，还有常用的数学符号和图形、大小写希腊字母和拉丁字母等内容；功能键区的内容除了标准键盘的功能键外，还应安排国标GB1988-80的C_0集和GB5261-85中的有关内容。关于这两个标准将在本书5·2节介绍。保留区又称机动区，是为进行盘外字处理和满足用户其它需要设置的专用区。根据盘外字处理的不同方法（见下段），该区设置有二、三百个字位的基本部件表等内容。

现举例示出4 K型和2 K型两种整字键盘盘面分区，如图3·26所示。图中(a)是4K型整字键盘的盘面。整个盘面有效工作面积有$53\times72=3816$个字位。其中，右下角的空白区域有$26\times17=442$个字位，放置非汉字字符和常用功能键。其余3374个字位是汉字区和保留区，可放置常用汉字及拼字用的基本部件(字根）。图中（b）是2K型整字键盘的盘面，共有$32\times64=2048$个字位。除右下角空白区域的$16\times20=320$字位为非汉字字符和功能键区外，其余1728个字位安排最常用汉字及拼字的笔画和部件（字根）。

各种型号的整字键盘盘面分区情况并不相同，在操作使用整字键盘以前，应先了解盘面的分区，才能做到正确操作使



(a) 4K型



(b) 2K型

图 3·26 汉字整字键盘盘面分区示意图

用，并进而逐步提高汉字输入速度。

② 盘面字表的汉字排列

整字键盘盘面字表的汉字排列设计，是一项技术性很强的工作。如同标准键盘的汉字输入编码一样，字表排列设计涉及语言文字学、数学、人机工程学和计算机技术等多种学科。字表中汉字的排列是否恰当，直接影响整字键盘的生产和使用，尤其对找字速度快慢的关系很大。

确定字表汉字的排列方式有两条基本原则：一是应便于找字和记忆，二是应使触笔的平均移动距离尽量小。其目的是减少找字检索的时间和减轻操作者的劳动强度，以提高整字键盘的汉字输入效率。

根据生产厂家的不同设计，现用的各种型号整字键盘字表排列方式是各不相同的。应用较多的有以汉语拼音音序字母分区排列、以部首笔形为序分区排列、以使用频度及词组联想为序分区排序等几种方式。

以音序排列的字表，主要面向懂得汉语拼音的操作人员。这种字表的排列规则是：在选择好所用汉字总数的基础上，以汉语拼音字母顺序，把字表划分成23个音区。每个音区内的同音节汉字按形旁相对集中，按照先上下、后左右的规则顺序排列在一个矩形区域内。汉字中常用的大写数字（壹、贰、叁……）和时间及日期用字（年、月、日、时、分、秒）应相对集中在字表右下方靠近非汉字区的内侧。

以形序排列的字表，主要面向不熟悉汉语拼音的操作人员。这种字表的排列规则是：在选择好所用汉字总数的基础上，以汉字起笔和第二笔笔形为序，以“横、竖、撇、点、折”分成25个形区。每个形区内的汉字按部首及笔画数相对集中排列，常用的大写数字和时间及日期用字的安排与音序排列的字表相同。

以使用频度为序排列的字表，主要是面向某些固定专业范

围的操作人员。这种字表是将最常用的高频汉字先排列在字表的中心位置，然后按照构词规律，将常用字互相搭配为常用词，逐层向外辐射状排列。

国家标准《汉字整字键盘盘面字排列》GB7513-87规定了以音序排列和以形序排列的两种字表，作为今后整字键盘盘面字排列的统一标准。该标准字表上的文字排列规则简单明确，盘面分配和字组划分充分考虑了提高检索找字速度的要求。采用这种标准字表的整字键盘，不仅能获得较高的汉字输入速度，还有利于整字键盘的推广应用。

国标整字键盘字表有2K型和4K型两类。现将国标2K型形序字表和2K型音序字表列于表3·12和表3·13，供读者参考。

③盘外字的输入

盘外字是指盘面字表中没有收集的汉字。由于整字键盘的盘面字量有限，不可能把各种情况下可能用到的汉字都收集进来；另一方面，由于汉字的使用频度相差极为悬殊，从实用角度考虑，也没有必要把大量非常用字都收集进来。实践证明，对于那些使用频度很低，不值得在字表占有“一席之地”的非常用字，采用“盘外字输入”的办法处理，是比较适当的。例如上面列举的2K型字表，由于盘面容字较少，操作时检字比较方便，但“盘外字输入”就更显得必要了。

为了输入盘外字，通常在盘面的功能键区内设有“盘外字输入”键。在进行整字输入操作时，遇到字表上没有的汉字，就首先按一下“盘外字输入”键，使设备转入盘外字输入状态，然后按照规定的盘外字输入方法进行输入。

盘外字输入方法类似于标准键盘的输入编码，有多种不同的方法。较常见的有拼字法、音形法、拼音法和代码法等数种。

表3·12

2K型键盘字表（形序）

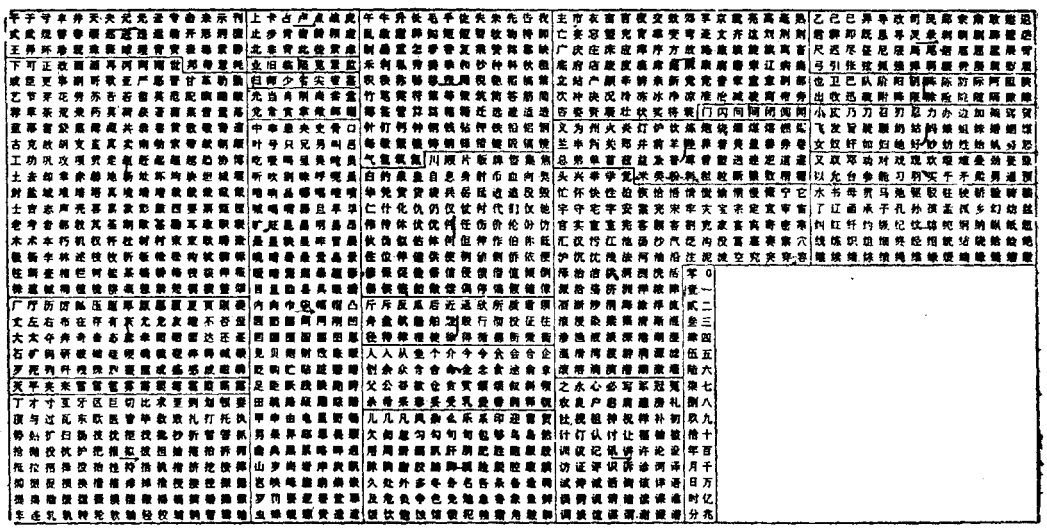
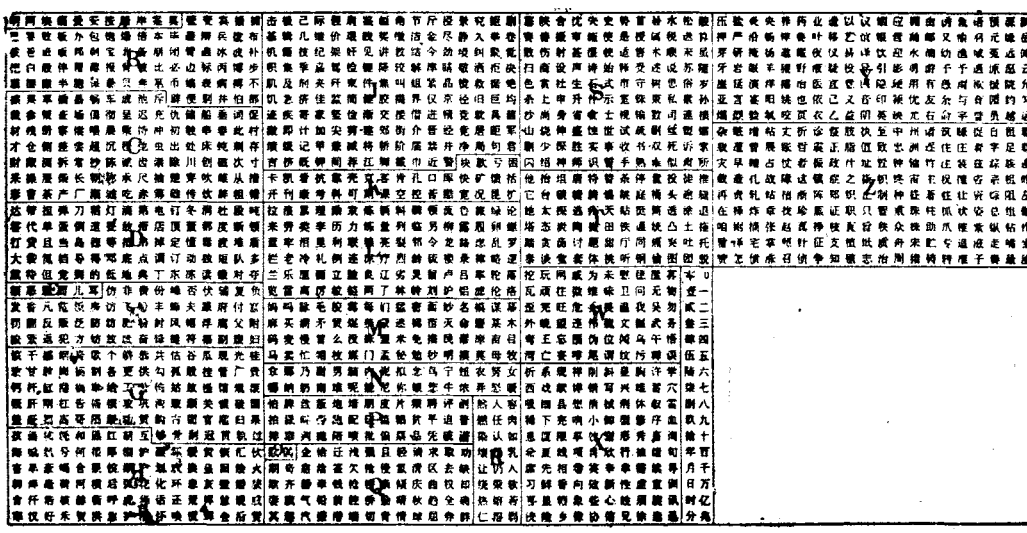


表3·13　2K型键盘字表（音序）



拼字法是利用字表上已有的单字、偏旁、部首或字根来拼出所需的盘外字。例如汉字“浒”在字表上没有，需要输入时，先按“盘外字输入”键，再依次输入部首“氵”和单字“许”，最后按“盘外字输入结束”键就行了。用拼字法输入盘外字，一般只需“两拼”或“三拼”即可，操作比较简便。

音形法和拼音法输入盘外字的基本方法与拼字法相似，只是所用的“原料”不同而已。但由于盘外字一般比较生僻，往往不知其读音，故音形法和拼音法的盘外字输入用得较少。

代码法是用国标码、区位码或者电报码的代码组来输入盘外字。它与汉字输入方案中的序号码法完全相同。用代码法输入盘外字虽然需要翻阅代码本或码表，但由于其对应性好，而且数量很少，故也常采用。

整字键盘盘外字输入方法目前尚未统一。因此，许多整字键盘都能提供两种以上的盘外字输入方法，以方便使用。

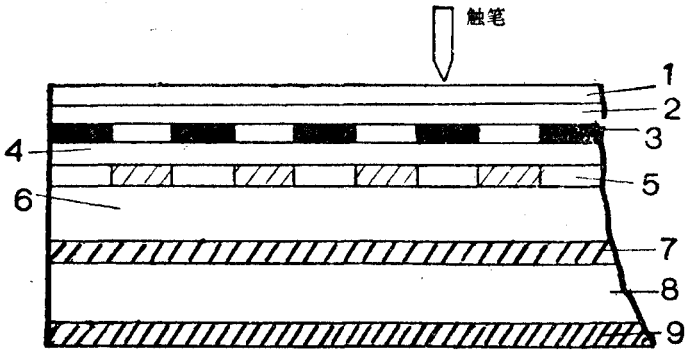
（3）笔触式整字键盘的基本工作原理

如前所述，全键式整字键盘和换位式整字键盘的基本工作原理与标准键盘是一样的，只是键位数量和布局的不同。至于笔触式整字键盘的基本工作原理则与上述几种键盘完全不同。目前，国内外生产的笔触式键盘种类很多，现以应用较多的静电耦合式、电磁感应式和压感式三种为例，简要讨论笔触式整字键盘的基本工作原理。

① 静电耦合式整字键盘

静电耦合式键盘是由屏蔽网络、X扫描导线组、Y扫描导线组和基板等部分，经精确定位后重叠起来组成的一个坐标定位盘。图3·27所示是一种国产的静电耦合式键盘的基本结构。

由图可见，该键盘主要部分分成三层。从下向上第一层是



1—保护层　2—文字表　3—文字感应块及屏蔽网络
4—基板2　5—Y扫描导线　6—绝缘层
7—X扫描导线　8—基板1　9—屏蔽层

图 3-27　静电耦合式键盘基本结构图

基板1，这是一块厚度约1.5mm的双面复铜板。基板下面有一层铜箔作为屏蔽层，用以屏蔽来自外部的高频辐射信号对字盘的干扰。基板上面布有若干根沿X轴向平行的X扫描导线。当由主控制器来的X扫描脉冲信号依次加在各条扫描线上时，这些导线就成为在坐标盘X轴向的脉冲信号发射体。

第二层是绝缘层，这是一块厚度约0.2mm的玻璃纤维布填料，起第一层和第三层之间的粘合和绝缘作用。

第三层是基板2，这是一块厚度约0.3mm的双面复铜板。该层下面布有若干根沿Y轴向平行的Y扫描导线。它的作用和原理与X扫描导线相同，是坐标盘Y轴向的扫描脉冲信号的发射体。

第三层上面是坐标点位置感应块和屏蔽网络。每个坐标点位置感应块都处于屏蔽网络的中心，并与下面X、Y扫描导线的交叉点一一对应。而盘面字表上的每个汉字字位又与每个位

置感应块一一对应。位置感应块是一块金属导体，它所感应到的X、Y扫描信号在块内平面上任何一点都是等电位的。当触笔触在感应块上的任意一点时，相应的X、Y扫描线上的扫描信号便经由感应块耦合到触笔内，并经过读出放大器送入主控制器。

主控制器是静电耦合式键盘的核心，它能把触笔送来的耦合信号转换成相应汉字或字符位置代码。现以一种国产的静电耦合式键盘为例来说明键盘控制器的基本工作原理。图3·28是它的原理框图。它是由图中除字盘(坐标盘)、触笔和输出缓冲器外的各个电路单元组成的。

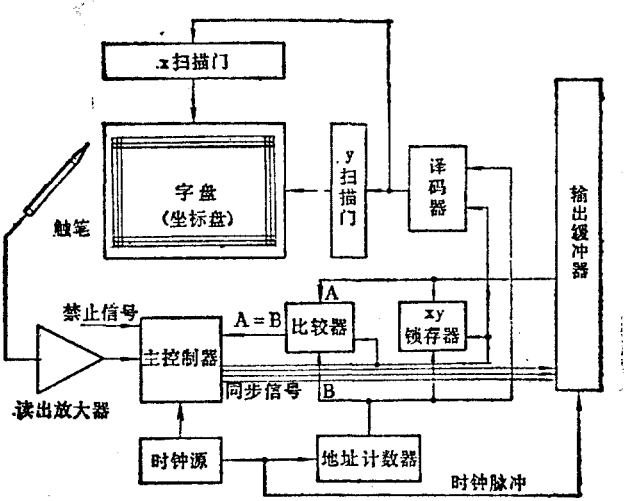


图 3·28 静电耦合式键盘控制原理框图

该键盘的坐标盘上，由正交的X、Y扫描线形成的48×64=3072个检测点，决定了盘面的最大容量是3072个字位。当键

盘在工作状态时，操作者将触笔触到盘面上的某一个字位，笔内的开关接通，使扫描计数器开始计数。通过译码器译码后，依次选通X扫描门和Y扫描门，并使X、Y扫描线都分别扫描两次。当扫描到对应于触笔触压点的X、Y扫描线时，通过触笔与扫描线的静电耦合作用，把扫描信号回授到主控制器。主控制器将先后得到两个X耦合信号和两个Y耦合信号。在得到第一个X（或Y）耦合信号时，主控制器将扫描计数器的状态X(或Y)在锁存器中锁存，以便与出现第二个耦合信号时的计数器状态相比较。比较符合时，扫描计数器的状态就是所触及字位的坐标值，即某一汉字（或字符）的位置代码。位置代码接收无误后，设备自动发出数据选通脉冲，将锁存器中的位置代码经输出缓冲器送到主机，然后通过软件方法（查表或简单运算程序）或硬件方法（专用的译码器或存储器）将位置代码转换成汉字或字符的标准代码。

② 电磁感应式键盘

电磁感应式键盘又称为磁笔触式键盘。它的基本工作原理是：当选字操作时，装在触笔内部的电磁线圈被接通电流，并产生较强的磁场。在字盘下面坐标盘内的读取线将磁场检出，从而得到触笔的位置信号。

图3·29是电磁感应式键盘坐标盘结构示意图。

如图所示，在坐标盘平面内布有55个X轴向、94个Y轴向的短路环，即读取线。在它们的正交处形成$55\times94=5170$个检测点。这些检测点与文字盘上的字位一一对应，成为总数5170个汉字、字符及操作功能键的笔触键盘盘面。当操作者用触笔触及盘面的某个字位时，笔内的微动开关接通，检测信号电流给电磁线圈励磁，产生的脉冲磁场使对应检测点的短路环中产生感应电势。回路导线上的感应电势方向，须视笔尖与回路的

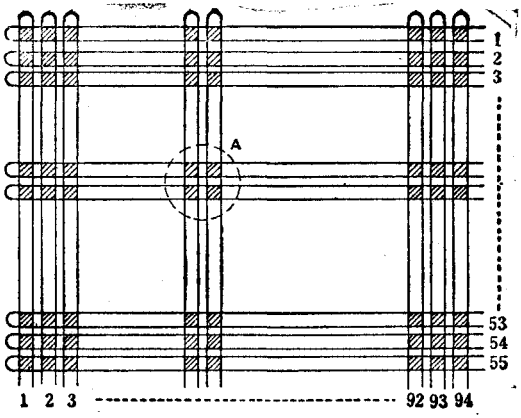


图 3·29 电磁感应式键盘坐标盘结构示意图

相对位置而异。以图3·29(b)为例（这是3·29(a)图中A点的放大图），设触笔笔尖在P_1点位置上，当笔内线圈产生脉冲磁场时，根据楞次定律，在触点P_1上方的导线X_1与下方的导线X_2上的感应电势方向是相反的，因此在X_1和X_2组成的回路中的总感

应势叠加，即$e_x = e_{x1} + e_{x2}$。对于导线X_3和X_4来讲，它们都在触点的同一侧，导线上的感应电势方向相同，且由于导线间距离很近，故感应电势的大小亦相差无几。因此，X_3和X_4导线组成的回路中无感应电势输出。同理，对Y轴向的短路环来讲，只有Y_1和Y_2组成的回路两端有感应电势$e_y = e_{y1} + e_{y2}$输出，其它的短路环回路两端均无感应电势输出。于是，控制器通过扫描读取X、Y两方向上的感应电势e_x和e_y，便可获得触笔触点位置的平面坐标值，即相应汉字和字符的位置代码。通过软件或硬件方法转换，便可得到该字的标准代码。

③ 压感式键盘

压感式键盘的基本构成如图3·30所示。

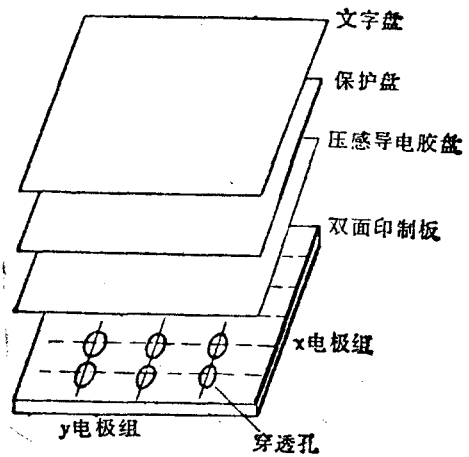


图 3·30 压感式键盘的结构示意图

图中最下层的双面印刷板上，是带有穿透孔的X、Y正交电极组。双面印刷板上叠一层压感导电胶盘。盘上加一层保护盘，最上面是文字盘。压感式键盘的基本工作原理是：当操作

者把触笔触压文字盘上某一字位时，压感导电胶盘的受压部位呈现导电状态，并经过盘下的穿透孔使X、Y电极组的某X电极和Y电极连通。此时，控制器在X电极组顺序加上检测电压，在Y电极组上顺序检测，便可测知触点的X、Y平面坐标值，即该字位的位置代码，然后便可获得相应汉字或字符的标准代码。

由此可知，压感式键盘的基本工作原理比较简单。由于触笔的作用只是施加压力，故不需特殊加工，也可用普通元珠笔或手指触压。而且坐标盘的结构也不复杂。压感式键盘的主要难点是需要选择性能优良的压感导电胶材料，使之具有明显的压力与电阻值变化特性，才能不致发生要求压力过大而造成漏输入或者过于敏感而造成错误输入的情况。

3. 专用键盘（中键盘）

专用键盘是指专门为汉字输入设计的非整字键盘。由于专用键盘的键位数量介于整字键盘与标准键盘之间，故又称为“中键盘”。

如前所述，用只有几十个键位的标准键盘输入汉字，需要先进行汉字编码，但设备是通用的，投资省；用有几千个键位的整字键盘输入汉字，一字一键，不需输入编码，但盘面字数多，找字困难，而且需要增加设备投资。为了解决上述矛盾，使两种键盘取长补短，于是就设计生产了汉字专用输入键盘。

专用键盘通常设置100～300个键位，因此不能用整字输入，仍然需要“拼字”输入。这种“拼字”法的实质也是汉字编码，但是由于键位比较多，所以编码规则远比标准键盘简单，一般只需二键（最多三键）就能输入一个汉字。而且专用键盘的键位设置比较灵活，编码方案的设计可以不受标准键盘几十个键位的限制。因而使汉字的输入操作更加直观，效率更高。

因此，在笔触式键盘普遍应用以前设计生产的汉字终端上，有不少都采用由生产厂家自行设计的专用键盘作为汉字输入设备。

专用键盘也有拼形输入和拼音输入两类。

拼形输入的专用键盘是以部件（包括成字部件，即单体字）设键的键盘装置。就是给每一个部件设置一个键位，再按照先左后右、先上后下，先外后内等简单的规则，将构字部件依次输入，便可达到输入汉字的目的。

前已提及，目前还没有划分汉字部件类型和数量的统一标准，根据不同的字形分解方法，构字部件的总数大约可分成五、六百个到一千多个不等。当然不可能把这些部件都单独设键，否则“中键盘”又变成“大键盘”了。因此专用键盘的设计者仍需按照不同部件的构字能力和实用频度，对它们进行筛选归并，使之减少到二、三百个以下，然后分类设键。

根据汉字字形分解的规律，选用的部件数越少，就得把字形分解越细；反之，增加选用的部件总数，可以降低对字形分解的要求。由于专用键盘的键位多，可以容纳较多的构字部件，这样可以简化标准键盘输入编码的字形分解规则，有利于操作者的学习和记忆，能明显提高汉字输入效率。例如有一种专用键盘，规定每个汉字输入时最多拆分成两个部件，也就是说，最多只需打两次键就能输入一个汉字。对高频字及可以作为单字的部件，则单独设键一次输入。

总之，拼形输入的专用键盘，既保留了整字键盘直观易学的特点；又具有码长较短，输入速度较快的长处。有一种能兼容多种编码方案的专用键盘，设有158个键位。其键面布局如图3•31所示。

拼音输入的专用键盘是一种以汉语拼音方案规定的声、韵

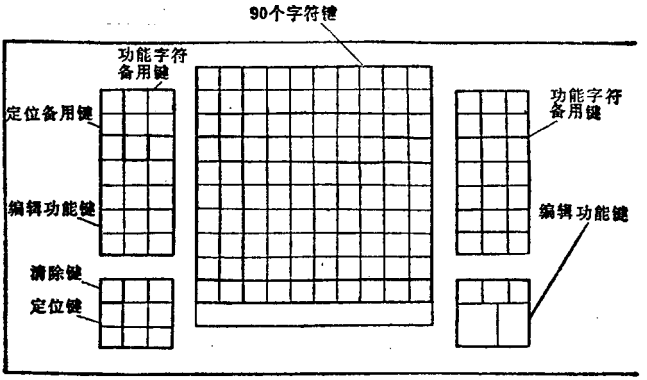


图 3·31 一种专用键盘的盘面布局图

母设键的键盘装置。如前所述，汉语拼音的声、韵母总数将近60个，但标准键盘的字母键键位只有26个。于是人们设计出各种一键多用的方案，达到了使用标准键盘以拼音输入汉字的目的。但是，这些拼音键盘由于受到标准键盘原有键位的约束，难以完全满足汉字拼音特点和手指快速打键的要求。因此又有人设计了一种能使音韵母单独设键的拼音（双拼）输入专用键盘。这种键盘的盘面布局可参看上节图3·11。

在现阶段，由于标准键盘输入编码方案的日趋完善，以及笔触式整字键盘检测技术的迅速发展，加上专用键盘的制造工艺难度大，产品的可靠性较差等方面的原因，使专用键盘的应用范围与标准键盘、笔触式整字键盘相比，近几年没有得到更大的发展。

3·5 汉字识别

汉字识别是用机器自动识别汉字字形的一种汉字输入方

式。

汉字识别的基本过程，是通过光学或压力的传感检测装置将字形变成电信号，应用计算机的智能特点对信号进行比较和判别，然后转换成汉字的标准代码输入。

汉字识别是计算机模式识别的一个重要分支，是第五代计算机(即智能计算机)视觉系统的组成部分。对于汉字终端和汉字信息处理系统来讲，汉字识别是一种高速、自动的先进输入方法，具有很高的实用价值和广阔的应用前景。

1. 汉字识别技术概述

（1）汉字识别技术的历史和现状

汉字识别是一门介于基础研究和应用研究之间的边缘技术学科，与人工智能、图象处理等学科有密切的交叉关系。汉字识别技术的理论比较复杂，研究的历史还不太长 。但进入80年代以来发展比较迅速，有些项目已经达到和接近实际应用的水平。

国外的字形识别研究始于50年代初期。1955年印刷体阿拉伯数字的光学字符阅读机（简称OCR，是 Optical Character Reader的缩写）问世以后，便大力开展对手写体字符识别的研究。在70年代初期实现了印刷体和手写体阿拉伯数字、拉丁字母和日语假名的OCR商品化。在此基础上进一步开展了对汉字字形识别的研究。1977年底，首次推出了可识别2000个印刷体汉字的OCR试验样机，可用于识别专利公报等技术资料。该样机的识别速度为每秒100个汉字，正确识别率为98.4%，拒识率为1.3%，误识率为0.3%，基本上达到实用水平。

我国于70年代初期开始进行印刷体和手写体字符识别的研究，主要是计算机用的字母、数字和符号识别，以及用于邮政信函自动分拣装置的手写体阿拉伯数字识别。在基本达到实用水平的基础上，于70年代中期开始对印刷体和有限制的手写体汉字识别研究，并已取得了积极的成果。在单一字体的印刷体汉字识别方面，对国标基本集的常用汉字（一级字3755个）正识率达到98.3%，识别方法已趋成熟，接近实用水平。在多种字体的印刷体汉字识别方面近年来取得了新进展，可以不用切换字典识别宋体、仿宋体和黑体三种一号字1800个，在采集样品不大于48×48点阵时，正识率可达95%。

手写体汉字识别的技术难度比印刷体汉字识别的难度更大。众所周知，手字体汉字的字形因人而异，变化多端。因此，就目前的科技水平来讲，以没有任何限制的手写汉字作为识别对象几乎是不可能实现的。即使是有限制性的手写体，字形变化仍然比较大，识别过程相当复杂。目前已经提出了几种基于汉字字体结构模式的识别方案，并取得了较好的实验性成果。

上面讲的印刷体和手写体汉字识别是非实时性的，就是应用计算机对已经印好或写好的汉字文稿进行识别。另一种手写体汉字识别是在联机状态下，应用计算机的实时处理能力，在操作者写字的过程中进行识别，称为手写体汉字实时识别。这种识别方法可以实时地抽取手写过程中的笔画长度、方向，笔画顺序（笔顺）和笔画数等特征参数信息，有利于对字形的判别。其中有些信息（如笔顺）在已写好或印刷好的字面上是无法取得的。而且，一般实时书写的字体都比较大，对检测设备的精度要求不太高，故成本比较低。近年来，国内外已有多种实时的有限手写体汉字识别装置问世，并已开始走向商品化。

（2）汉字识别的基本理论方法

前已提及，汉字识别是计算机模式识别的一种。所谓“模式识别”，就是用机器直接识别和理解声音、文字、图象等人们所熟悉的模式（模拟）信息。模式识别的基本理论方法主要有两大类，一类称为决策理论方法，另一类称为句法方法。

决策理论方法又称为统计方法。这是发展较早、也比较成熟的一种模式识别方法。决策理论方法的基本过程是：首先将识别对象（即模式信息）数字化，成为适宜于计算机处理的数字信息。然后将数字信息进行预处理，用于除去混进来的干扰信息，并减少某些变形和失真。预处理以后进行特征抽取，从输入信息中抽取一组特征，以便于进行分类和判定，达到正确识别的目的。

在决策理论方法中，特征抽取占有重要地位。因为在识别对象进行数字化时，一个模式常常要用很大的信息量表示，其中必含有很多冗余信息。所谓“特征抽取”，就是要选定一种度量，使它能够代表输入模式，而使其中的冗余信息尽可能减少。这个度量，通常可以用一个特征矢量来表示。由特征矢量可计算出相应于输入模式各类别的判别函数(或称鉴别函数)。通过对判别函数的比较，就可实现分类和判别。

句法方法又称为结构方法或语言学方法。这是近几年得到迅速发展的一种模式识别方法。

句法方法的基本原理是：先把一个模式描述为较简单的子模式的组合；子模式又可描述为更简单的子模式的组合。这样下去，最终可得到一个树形的结构描述。在底层的最简单的子模式叫做模式基元。把模式用一组基元以及它们的组合关系来描述，称为模式描述语句。基元组合成模式的规则由语法指

定。这相当于在语言学中句子是用词来组合、词用单字来组合，它们的组合关系则由语法指定一样。

在句法方法中选取基元，相当于在决策理论方法中抽取特征。但选取的基元数量一般比特征参量的数值少。通常要求所选的基元既能对模式提供一个紧凑的、反映其结构关系的描述，又要易于抽取。而基元本身不应含有重要的结构信息。一旦基元被鉴别，便可通过句法分析进行识别。即分析给定的模式语句是否符合指定的语法。符合某类语法的，就被分入该类。

在汉字识别中，决策理论方法已经广泛应用于印刷体汉字识别方面，句法方法则主要用于有限手写体汉字识别方面。当然，这两种方法不能截然分开。在许多应用场合，往往将它们结合起来，分别应用于不同的层次。例如，当使用句法方法进行手写体汉字识别时，基元的选取可用决策理论方法来实现，这样可以收到较好的效果。

以下简要介绍两种比较成熟的汉字识别技术：印刷体汉字识别和手写体汉字实时识别。

2. 印刷体汉字识别

印刷体汉字识别应用光电扫描技术和计算机的判别处理能力，高速、自动地识别和读取印刷在纸上的汉字信息。印刷体汉字识别的应用，能把大量已有的、而且与日俱增的文献资料输入计算机信息处理系统，这对于建立各种中文资料档案数据库具有十分重要的意义。

印刷体汉字识别的基本工作流程如图3-32所示。

由图可见，汉字识别的第一步是将被识别的印刷体汉字文稿进行检测，就是通过光电扫描将字形变换成适合机器处理的

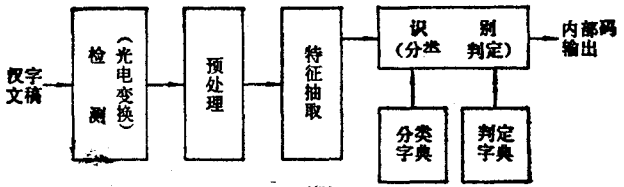


图 3·32 印刷体汉字识别基本流程

数字信号。第二步需要将检测到的数字信号进行预处理，预处理的内容是：去除由于纸面污点、空白等造成的干扰信号；根据一定的准则压缩冗余信息；对文字的大小、位置和笔道粗细进行规范化。第三步是特征抽取，就是从经过预处理的规范化信号中抽取反映字形某种性质的图形特征，例如字形的笔画长度，周围笔画密度，端点和交叉点个数等，并以一种称为判别函数的多维向量形式表示。最后把判别函数送到识别部分进行分类和判定，在分类字典和判定字典中存放着以多维向量形式表示的标准字形，把未知字形与标准字形进行匹配、比较，选取二者最接近的标准字形，就作为识别的结果输出。

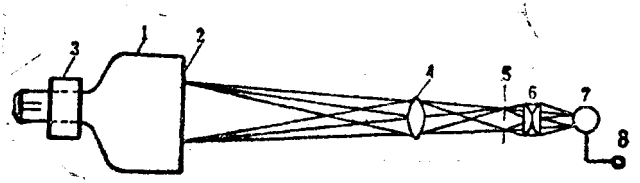
（1）字形检测设备的基本原理

字形检测设备是应用某些器件的光电变换特性，通过扫描把字形信息转换成电信号的装置。字形检测设备的种类比较多，按照它们所用的不同器件，其工作原理和特性各有差别。常见的有飞点扫描器、光敏二极管矩阵，激光扫描器和电荷耦合器件（CCD）等。

① 飞点扫描器

飞点扫描式字形检测器（简称飞点扫描器）是利用集中光能源来扫描被识别字形，以获得电信号输出的一种传感装置。

飞点扫描器的基本结构原理如图3·33所示。



1—投射管　2—投射管屏　3—偏转线圈
4—透镜　5—被识别文稿　6—聚光镜
7—光电倍增器　8—二值图形输出

图 3·33　飞点扫描器结构示意图

由图可见，飞点扫描器是由投射管、偏转系统、透镜、聚光镜、光电倍增管等部分构成。投射管 1 是一个亮度高、聚焦性能好的小型显象管，作为飞点扫描器的光源。工作时，投射管屏 2 发出一个集中的光点，照射于被识别文稿 5 上，每一瞬间只照射其中的一个象素。被照亮的象素发出的反射光或透射光（图中为透射光），经过聚光镜 6，射入光电管或者光电倍增器 7 的阴极上，按照透射光的亮度大小、转换成相应的电信号。

偏转线圈 3 的作用，是使投射管阴极发出的电子束在管屏 2 上扫描，从而使投射管管屏发出的亮光在识别文稿上进行扫描。“飞点扫描”即由此得名。

在光点顺序扫描文稿的过程中，如果管屏的余辉时间不大，可以认为在每一瞬间只有一个被扫象素在透光，其透光强度取决于该象素的浓度（或透明度）。有字迹处的象素浓度大，透光强度较弱；空白处的象素透明度好，透光强度较强，于是光电倍增器就产生相应的被调制信号。由于“飞点”的瞬时亮度很大，光能利用率很高，从这种装置获得的信号能量较

Retro Workshop
果粉工作室

大，从而可以得到高质量的图形信号。

飞点扫描器的一大特点是扫描速度很快，输入文稿的检测速度主要受制约于管屏荧光材料的余辉时间。假设判别一个汉字需检测100×100个“点”（通常只需40×40个点），如果余辉时间缩短到0.1微秒，则飞点扫描速度可达每秒1000个汉字。

飞点扫描器的另一特点是扫描控制方式比较灵活。由于投射管的光点直接受偏转线圈控制，通过改变偏转电路参数便可随时改变飞点扫描方式。因此，在字形识别系统中应用较广。

飞点扫描器的缺点是，如果受到外界光的影响，会降低信噪比。因此必须采取遮蔽外界光的措施。

② 光敏二极管矩阵式传感器

光敏二极管矩阵式传感器，是利用光敏二极管的光电变换性能，将光源照射到文稿上的反射光或透射光，经聚光镜投射到光敏元件上产生相应的电信号，然后顺序地提取这些信号。其工作原理如图3·34所示。

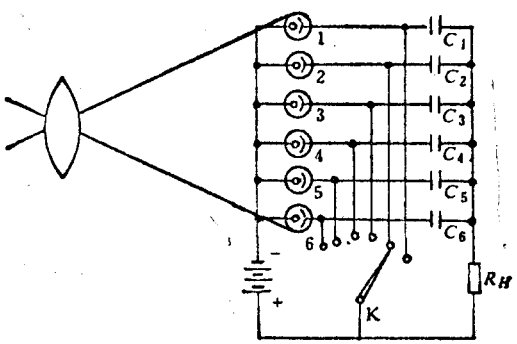


图 3·34 光敏二极管传感器工作原理示意图

图中，光敏二极管1至6，是二极管矩阵一列中的几个（其它各列均未画出），它们的阳极都连接在一起，而阴极互

相绝缘，并通过一个单独的电容器与负载电阻R_H相连。在光照下，各个光敏二极管中流过电流的大小，与投射到该管上的象素亮度成正比。这个光电流对相应的电容器充电，充到正比于光电管照度的电位。使转换开关K不停地切换，各个电容器（C_1至C_6）顺序通过负载电阻R_H放电，在R_H上得到与光敏二极管照度成正比例的电信号。

光敏二极管矩阵传感器是一种经济有效地利用光能的检测装置。因为它通过向电容器充电积累光产生的能量。由于，在每一象素上所用的光能，不只是在被扫描瞬间的极小光量，而是在传送整幅文稿时间内所获得的光能，所以能获得较强的信号输出。这就是信号转换和传送中广泛采用的积电原理。

③ 激光扫描式检测器

激光扫描式字形检测器是利用激光器产生的光束来扫描文稿的检测装置。由于激光光源纯度高、亮度大、方向性好，可使扫描光点凝聚得很小，因而大大提高文稿识别的分辨率。同时，入射光能强，可以减少对外界光的屏蔽要求，提高检测器输出信号的信噪比。

激光扫描式检测器的基本工作原理如图3-35所示。

由图可见，激光源是通过垂直反射镜进行光束扫描的。垂直反射镜是一种圆柱体状的多面镜鼓，圆柱体的侧面切成很多镜面，镜上镀有高反射率的金属膜，具有极好的反射性能。

进行字形检测时，激光光束被转动的多面镜鼓反射，通过透镜照亮文稿上的各个象素，并依次逐行扫描。被照象素发出的反射光或透过的透射光被光电倍增管接收，并产生相应的被调制信号输出。

由于采用机械转动机构，激光扫描器的扫描速度较低。为了获得稳定的光栅图象，镜鼓尺寸、形状、转速和文稿移动速

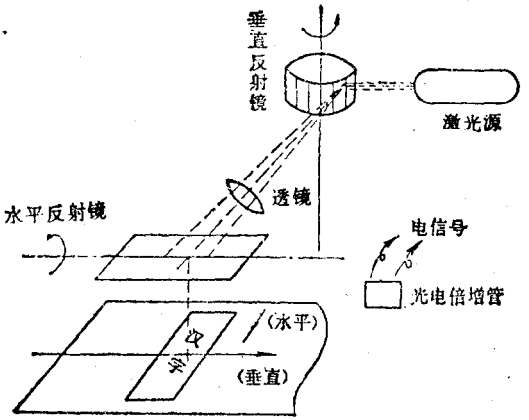


图 3·35 激光扫描式检测器原理示意图

度等参数都必须严格配合好。

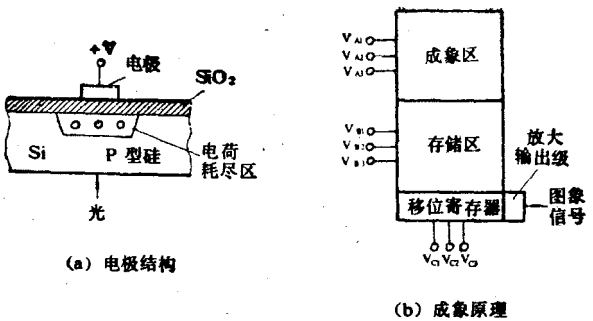
④ 电荷耦合式字形检测器

电荷耦合式字形检测器，简称电荷耦合器（Charge Coupled Device，简写为CCD)。由于它具有一些独特的性能，故发展极为迅速，已经在字形检测、摄象和信号存储等方面得到了广泛的应用。

电荷耦合器是一种MOS集成电路。实质上，它是由时钟脉冲电压来产生和控制半导体势阱的变化、以势阱内的电荷量来表示不同光强状态的动态移位寄存器。

图3·36（a）是CCD一个电极的基本结构示意图。

如图所示，当CCD的电极被加上适当的偏压时，该电压形成的电场穿过SiO_2薄层，在电极下方形成一个电荷耗尽区，而在SiO_2—Si界面上得到一个储有少数载流子的势阱。外加的偏压愈大，势阱就愈深。将许多电极排成阵列，并和适当的输入、输出电路相连接，就成了基本CCD。



. 图 3·36 CCD的电极结构及电荷转移原理图

当硅晶体受到图形象素光的投射而产生电子空穴对后，少数载流子在电场作用下被吸入较深的势阱中，投射光越强，势阱捕获的电荷量就越多。于是，图形的光象就转变成由积累电荷量描绘的电子映象。这些电荷储存在CCD中，完成了光电转换及储存信息的过程。

图3·36(b)是CCD分区成象、储存和转移的结构示意图。

当电荷积累过程结束，成象区形成电荷图象后，在外加驱动时钟脉冲作用下，电荷图象迅速转移到存储区。然后成象区又开始新的电荷积累过程。存储区的内部结构与成象区相同，但表层用一层不透光的金属复盖着。存储区后面连接输出移位寄存器及输出放大器。在同步脉冲电压的作用下，存储区的电荷逐行转移到移位寄存器，然后放大输出连续的图象信号。

（2）字形信息的预处理

字形信息的预处理，就是对光电检测器输出的字形信息进行特定的标准化处理。预处理的主要内容是：实现字形的规格化、除去噪音或污染、笔画的细化等。

① 字形规格化

字形规格化是把输入的字形变换到识别系统要求的大小或位置，是预处理的第一项内容。如果输入的字形大小不符合要求，可把该字所在的二维平面纵横两方向分别乘以适当的比例因子，便可将字形放大或缩小。如果由于文稿印刷质量不好或检测过程中扫描的非线性造成的字形变形或位置偏移时，应将文字所在区域的中心对准特定的坐标位置，再加以适当移动或旋转，便可纠正过来。

② 除去噪音或污染

预处理的第二项内容是除去噪音或污染。在光学扫描系统对纸面上的字形信息扫描检测时，由于纸张质量不好或印刷过程中的污染等原因，会造成对字形信息的噪音或干扰。例如在空白的区域中有了黑点，或者笔画线条中有断线和字迹不清等现象。对于文稿上孤立的黑点或白点的干扰，可以用下法除去。

图3·37是去除孤立点杂音的示意图。如果在一个很小的区域里，有一个孤立的白点P，它周围的8个点都是黑点，则可把P点看成是印刷中缺漏带来的孤立点杂音，可以用下列算法将它除去。

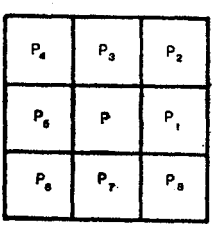


图 3·37 去除杂音示意图

设白色象素点信号为0，黑色象素点信号为1，则当 $P_1 \wedge P_2 \wedge P_3 \wedge P_4 \wedge P_5 \wedge P_6 \wedge P_7 \wedge P_8 = 1$ 成立时，（式中∧

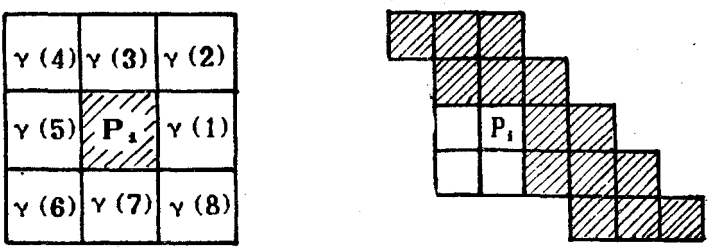
为逻辑与）

令　P = 1

同理，可以用相似的办法除去孤立的黑点，如此继续进行下去，便可除去纸面孤立污染点带来的噪音干扰。

③　字形笔画的细化

所谓字形笔画的细化，就是将文字图形的笔画处理成为只有 1 比特线宽的笔道。如图3·38所示，是一种称为“井罩法”的笔画细化原理示意图。



(a) 井字罩　　(b) 细化

图 3·38　井罩法细化原理图

图中（a）是一个井字罩，由 9 个网点组成。中心点与字形中的某点P_i重合，来检测它周围 8 个点的状况，用γ(k) 表示它们的值（k = 1，2 … 8）。例如井字罩右上角（2 点）有笔道通过，则γ(2) = 1，表示该点是黑点。如左下角（6点）没有笔道通过，则γ(6) = 0，表示该点是白点。且设γ(9) = γ(1)。P_i点的值$C(P_i)$用下式定义：

$$C(P_i) = \sum_{k=1}^{8} |\gamma(k+1) - \gamma(k)|$$

当满足下面条件时，应删去P_i点

(a) $C(P_i) = 0$ 或 2

(b) $\sum_{k=1}^{8} \gamma(k) \neq 1$

(c) $\gamma(1) \wedge \gamma(3) \wedge \gamma(5) = 0$

(d) $\gamma(1) \wedge \gamma(3) \wedge \gamma(7) = 0$

式中∧为逻辑与。

图中（b）的P_1点满足上述条件，故该点(黑点)应被删去。如此继续进行下去，可以得到笔道宽度为1比特的细化字形。

井罩法还可以用于字形的特征抽取，检测出笔道的端点、三叉点、四叉点等特征点，检测相应的笔道长度和方向等特征，在手写体字形识别中有广泛的应用。

（3）输入字形的特征抽取

字形的特征抽取，是将经过预处理的输入字形信息转换成特征空间的一个多维向量（判别函数），以便与分类字典中的标准字形向量进行匹配比较，实现分类和识别。

前已提及，特征抽取在模式识别的决策理论方法中具有重要的作用，但目前尚无成熟的理论指导，只能通过对识别对象的具体分析来选择。印刷体汉字识别的特征抽取方法有多种，常用的有复杂指数法、四周代码法、结构分析法等。

① 复杂指数法

复杂指数是指字形结构的单位扩展量中的笔画长度，是文字笔画复杂程度的一种量度。复杂指数有整体复杂指数、横向复杂指数和纵向复杂指数三种。

众所周知，汉字字形繁简不一，笔画数相差很大。最简单的字形只有一笔，复杂的字形有二、三十笔。显然，字形的复

杂指数特征同人们的直观感觉是一致的，其物理意义比较明白："一"和"翼"相比，它们的整体复杂指数差别极大；"量"和"绷"相比，"量"字的纵向复杂指数比较大，"绷"字的横向复杂指数比较大；"一"的纵向复杂指数最小，而"1"的横向复杂指数最小。"画、影、联、姬"等字的横向复杂指数和纵向复杂指数比较接近。现举一些常用汉字为例，它们的复杂指数分布如图3-39所示。

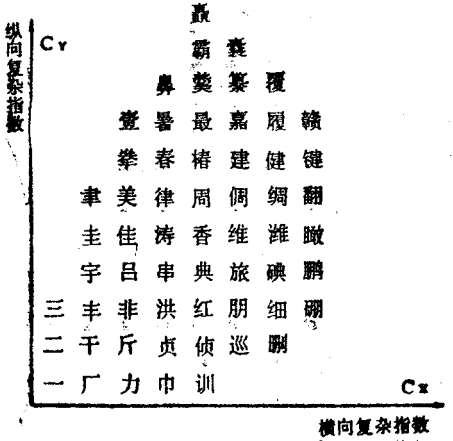


图 3·39　字形复杂指数分布示意图

由于字形的复杂指数不会因为文字位置的偏移而发生变化，因此，用复杂指数法抽取的字形特征量，对文字位置偏移的抗干扰能力比较强。

② 四边代码法

四边代码法是着眼于抽取字形四周的笔画分布特征来实现文字分类的方法。

每一个汉字都可看作是一个方块图形，具有上、下、左、右四条边。我们根据文字的线宽选择一个适当的宽度，用这个

宽度在文字图形的四条边线内侧分割出四个矩形小区域。然后，按照每个小区域中含有文字笔画的多少，选择适当的阈值来划分等级，并赋以相应代码。这样，每个汉字就可用四位数字来表示，这就是该字的四边代码。现以“中、国、字、用”四字为例，来说明四边代码的取码方法。如图3·40所示。

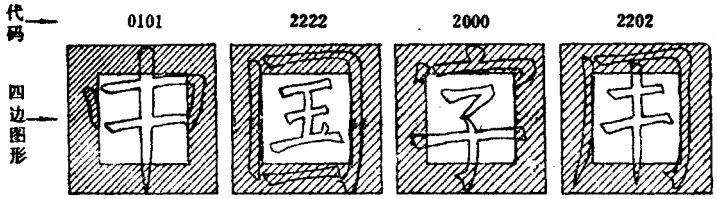


图 3·40　四边代码取码示意图

先从上边开始，按顺时针方向在第一、二、三、四边的内侧，划出适当宽度的四个小区域。然后以该区内笔画浓度为准，分成0、1、2三个等级，则可取出这四个字的四边代码值为：中（0101）、国（2222）、字（2000）、用（2202）。

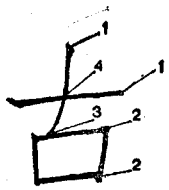
可以看出，汉字字形的四边代码特征，对文字位置偏移的变化很敏感，但对笔画中的断线、污点等有很强的抗干扰能力。

③　结构分析法

结构分析法是着眼于抽取字形结构特征的方法。表示文字结构特征的量有多种，如笔画的起点、长短、方向、分布、交叉等，我们可以选取其中的一种或几种特征量，作为对字形分类判别的依据。这里介绍一种文字结构特征点的基本概念。

表示字形结构特征点有端点、二分支点（屈折点）、三分支点（顶交点）和四分支点（交叉点）等，如图3·41所示。

在抽取字形结构特征点时，先用井罩法或其它方法，将文



1— 端点　　2— 屈折点
3—顶交点　　4— 交叉点

图 3·41　字形结构的特征点

字笔画细化为1比特宽度的笔道（参见 P.156③字形笔画的细化）。然后，用类似去杂音的办法（参见 P.155②除去噪音或污染），来确定文字图形中每一个交叉点（P）的特征。再通过逐点计算判断，便可确定该字各个特征点。

文字结构分析法中还有对笔画走向、长度及分布的特征分析等内容，这里就不一一列举了。

文字结构特征抽取受字形位置偏移和断线污点等干扰的影响比较小，有较强的抗干扰能力。但特征抽取过程比较复杂，主要应用在手写体汉字（书面文字）的非实时识别方面。在印刷体汉字识别中，主要用复杂指数法和四边代码法来进行分类。

（4）输入字形的分类和判定

对输入字形进行分类和判定，是字形识别的最后一步，其主要内容是用抽取的输入字形特征向量与字典中的标准字形进行匹配比较，把比较最接近的字形识别为输入字形，送入主机进行处理。

在西文文字识别系统中，作为识别对象的字母、数字及少量图形符号，总数一般在100个左右。因此，输入字形与字典中标准字形的匹配比较可以一次完成。但在汉字识别系统中，由于待识别的汉字字形结构复杂，字数很多。如果采用西文识别系统的一次比较办法进行识别，不但占用的内存量很大，而且匹配比较过程的矩阵运算十分复杂，所需的时间较长。随着识别数字的增加，将使系统的识别效率明显降低。

因此，印刷体汉字识别的匹配比较要分两步进行：第一步先对输入的未知字形进行分类，使每类中的预选字数限制在100个以内。第二步再将未知文字与相应类别中的预选字形进行匹配比较，最后得出识别结果。这样使识别系统的效率大为提高，方能满足实用要求。

①分　类

分类是指确定输入未知字形的大类别，也叫作粗分类。分类的基本方法，是把输入未知字形的特征向量与分类字典中的各类标准字形向量进行比较，看它最接近那一类。为此，就要在各类之间确定一个分界线，以便于确定它的归属类别，这个“分界线”就是判别函数。

假设输入未知字形x的特征向量为X，分类字典中有m类标准字形向量ω_1、ω_2……ω_m。与此对应，就有m个判别函数$G_1(X)$、$G_2(X)$……$G_m(X)$。

根据判别函数进行分类的算法是：若$G_i(X)>G_j(X)$成立，则x属于ω_i类；反之，x就属于ω_j类。其中i和j是1与m之间的一个整数。若$G_i(X)=G_j(X)$则x是在ω_i和ω_j两类的分界线上，无法分类，还必须引入其它特征重新分类。

判别函数的选择不是唯一的，可以根据抽取的不同特征向量进行选择，并据此做出各种不同的分类器。应用最广泛的一

种判别函数是线性判别函数，它是未知字形x的所有特征向量的线性组合。

一种能识别2000个字形的印刷体汉字识别装置，使用复杂指数作为分类特征时的平均每类预选字数为190个；使用四边代码作为分类特征时的平均每类预选字数为270个。两者同时使用时的每类预选字数可以减少到100个以下，最少可减到50个左右，使判定过程得到简化。

②判　定

字形识别判定是在分类的基础上，将输入的未知文字与同类别中的标准字形进行比较匹配，以得出输入文字的最后判定结果。

在印刷体汉字字形识别中，最简单的判定方法是模板匹配法（或称为样板匹配法）。就是把未知文字的特征向量X与存放在判定字典中的标准字形模板特征向量μ相比，计算它们是否相同或者相似。

设有两个标准字形A和B的向量μ_A和μ_B，要判定未知字形x是A或B，就要比较特征向量X是类似μ_A，还是类似μ_B。比较判别的具体方法很多，这里介绍两种方法：一种是距离判别法，一种是相似度法。

距离判别法是通过计算向量X和μ_A或μ_B之间的距离大小来确定归属。根据任意两点x、y之间的距离公式：

$$d(x,y)=\left[\sum_{i=1}^{n}(x_i-y_i)^2\right]^{\frac{1}{2}}$$

其判别法则为：

当　$d(X,\mu_A)<d(X,\mu_B)$时，则$x=A$

当　$d(X,\mu_A)>d(X,\mu_B)$时，则$x=B$

相似度判别法是用一个相似度函数S(X,X′)来比较两个向量X和X′。其值较大时，两向量相似。若以两个向量之间的夹角作为相似性的度量，称为单纯相似度。可用下列简单公式表示单纯相似度函数S(X,X′)：

$$S(X,X') = (X,X') / |X| \cdot |X'|$$

式中，(X,X′)为两向量的内积，|X|和|X′|为各自的模。可以看出，单纯相似度函数S(X,X′)就是两向量X和X′夹角的余弦。当夹角为0时，S(X,X′)＝1，表明两向量相合，否则，S(X,X′)的值大于0，但总小于1。因此，可以用S值的大小来判定：使S值最接近1的X′，就是未知字形向量X的判定结果。

单纯相似度法是一种识别字母、数字的有效方法，在印刷体汉字识别系统中也被作为一种最基本的方法。但是，在实用中印刷的文字图形会混有各种干扰斑点所形成的噪音，而且未知文字的图形位置也很难与标准图形完全一致。这些情况，都会造成单纯相似度法识别质量的严重下降。为克服这个缺点，引入了复合相似度和混合相似度判别法。

采用复合相似度法对混有噪音的文字图形能稳定地识别，对位置偏移也有较强的抗干扰能力，但对非常相似的文字，如王—玉、哀—衷等仍不能准确区别。混合相似度法是在复合相似度法的基础上，再进一步突出相似文字间的差别部分。这种方法的识别力较强，对于有若干个相似字形的情况，也可作出最终的判定。

3. 手写体汉字实时识别

（1）概述

手写体汉字实时识别，是在“联机”状态下，利用计算机的实时处理能力，边写字边识别，实现汉字输入。由于它使用

较方便，设备较简单，可以发展成一种普及和廉价的汉字输入手段。

前已提及，就目前的技术水平来讲，不论是实时或非实时的手写体汉字识别，都对操作者写字时的书写笔画规则有一定的要求。不加限制的连写和草书汉字的识别是几乎难以实现的。因此，在“手写体汉字”之前通常需冠以“有限制”几个字。当然，设备要求的限制条件，应该在一般书写习惯能够接受的范围之内。否则，过多过严的限制条件就会降低这种输入手段的实用价值，甚至使大多数人无法接受，影响推广应用。就目前已达到的设备水平来讲，对书写者的限制条件通常有以下几条：

① 以楷书字形的笔画为基础，一笔一画要分清，不能连笔。

② 基本笔画的运笔方向应该在限定的范围以内。就是要“横平、竖直”，撇、捺大致在左下和右下方向，不能过于偏离。

③ 应达到一定的笔画长度，使“撇、捺”与“点”有明确的界限；应按习惯笔顺书写。

④ 整个字形应写在规定的方格区域内；偏旁应与字体其余部分分开，不能粘连在一块。

当然，这里是指一般的规定，并非所有的手写体汉字识别输入装置都有这些要求。有的输入装置对笔顺不作要求，有的甚至可允许个别连笔笔画。应当指出，对手写字的限制条件多少，是衡量手写体汉字识别系统性能优劣的重要指标。

手写体汉字实时识别的基本流程如图3·42所示。

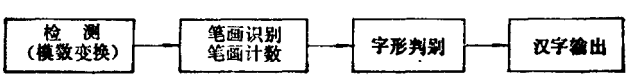


图 3·42 手写体汉字实时识别的基本流程

（2）检测

手写体汉字实时识别的检测设备是一个简单的模/数变换装置。它主要是一块边长约100～200mm的方形书写板和一支书写笔。工作时，设备不断地检测笔在书写板上写字时的位移量，把它变换成数据信息输入。

书写板通常采用平面型的图形输入板，在有效书写面积内排列着N×N个点阵。检测书写笔位移量的物理过程与上节介绍的笔触式键盘基本相同。但书写板的X、Y坐标的分辨率比较高，约为每毫米4～10线。当操作者按照规定在书写板上写字时，在系统实时处理功能支持下，能即时地抽取以下信息：

① 笔头在面板上移动轨迹的坐标值。

② 根据笔头接触和离开面板的信息，区分出字形的笔画数和笔画顺序。

③ 写完一个字的结束信号。

由于人手写字的速度比设备的检测运算速度慢得多，所以在书写过程中设备能根据上述信息立即计算出每一笔画的端点位置、笔画长度、笔向、笔形等参数，并及时储存起来。同时还应禁止重复数据输入。

（3）笔形识别

笔形识别是设备在抽取了笔画移动轨迹坐标值以后，通过一定的逻辑运算将它们变换为构成汉字字形的各种基本笔画。现举例说明识别“点、横、竖、撇、捺”五种基本笔画的逻辑运算方法。

假定以书写板的左下角为X、Y平面直角坐标的原点。任意一个单向（不拐弯）笔画的起点坐标为（X_1，Y_1），终点坐

标为 (X_2, Y_2)。规定上述五种基本笔画在X轴和Y轴上投影长度阀值（即界限常数）为C，则它们的逻辑判别式为：

“点”：$(|X_1 - X_2| \leqslant C) \wedge (|Y_1 - Y_2| \leqslant C)$成立；

“横”：$(|X_1 - X_2| > C) \wedge (|Y_1 - Y_2| \leqslant C)$成立；

“竖”：$(|X_1 - X_2| \leqslant C) \wedge (|Y_1 - Y_2| > C)$成立；

“撇”：$(|X_1 - X_2| > C) \wedge (|Y_1 - Y_2| > C) \wedge (X_2 < X_1) \wedge (Y_2 > Y_1)$成立；

“捺”：$(|X_1 - X_2| > C) \wedge (|Y_1 - Y_2| > C) \wedge (X_2 > X_1) \wedge (Y_2 < Y_1)$成立。

式中∧为逻辑与。

我们以“0、1、2、3、4”分别代表“点、横、竖、捺、撇”，可画出上列各式的逻辑关系图如图3·43所示。

（4）字形判定

如前所述，任何一个汉字字形都是一组基本笔画的有序组合。因此，在笔形识别的基础上，再加上它们之间相对位置及结构特点，便可判定输入字形。

字形判定的基本方法是把常用字的标准笔画有序组合做成字典预先存储在机内。在系统识别了各种输入笔画及其它有关参数以后，通过与标准字形字典进行比较，取最接近或相同的标准字形判定为输入字形。对于笔画和笔顺相同的“重码”汉字，则在字典中还要存储它们各自的笔画端点坐标值，以便对输入字形的有关参数作进一步的比较，就可以区分重码字。

例如“井”和“开”，两个字形的笔画数、笔形、笔顺都相同，则需要读取它们第三笔或第四笔的起点Y坐标值。如果该值大于第一笔的起点和终点Y坐标值，再与字典中已存的相应参数加以比较，便可判定输入字形是“井”还是“开”了。

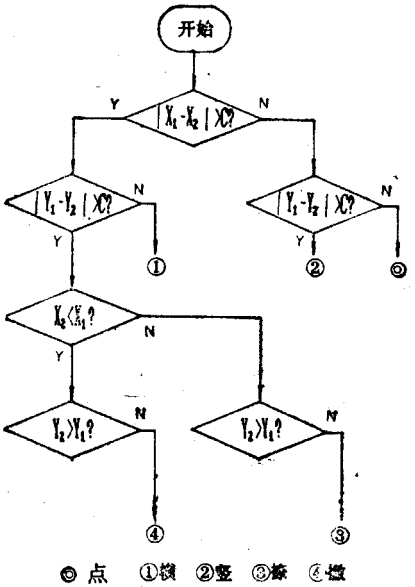


图 3·43 简单笔形识别的逻辑关系图

还有一种字形判定方法，是在笔形识别之后，再增加判别字形的偏旁，并且可以得到系统实时处理的支持，使判定较易实现。

此法的基本原理是先将汉字字形粗分成左右型和非左右型两大类。可以根据先写出的、位于书写板左边的几个笔画位置，判定出左偏旁的类型。然后对其余部分的笔画进行计数和识别比较。这样可以节省字典的存储量，并且提高识别速度。

在多数实时识别系统中，写字时的笔顺是作为一个动态特征参数，抽出来与标准字典进行比较判别的。但是许多常用字

的笔顺也往往各不相同。为了减少由于笔顺不一致带来的麻烦，有一种不依赖笔顺信息的实时识别装置，它的基本原理是：当把笔画轨迹坐标取样以后，用计算距离的公式算出标准字形笔画与输入字形相应笔画之间的距离。把笔画间的距离之和作为文字间的距离，于是可以取出文字间距离最小的标准字形作为判定的输入字形。

此外，正在发展中的应用模糊集图形识别法汉字识别系统，可以在笔画数略有不准确的情况下（即书写过程中的个别连笔），也能达到对字形的正确识别。这标志着手写体汉字识别正迈向实用化的新阶段。

四、汉字存储及输出技术

对于汉字终端和常用的汉字微型机来讲，汉字存储技术和输出技术的主要内容，就是如何实现把存于机器内部的汉字标准代码（内部码）变成汉字字形显示或打印出来。

汉字字形的输出，离不开存储字形信息的字形发生器，或称为汉字库。要把抽象的“汉字代码”变成眼睛看得见的“汉字字形”，必须从字形发生器中读出汉字的字形数据，送到显示器或打印机，才能实现。

把汉字信息转换为语音，也需要一个存储语音参数的存储器。通过语音合成器将语音参数合成所需的语音信号，并由扬声器输出。图4·1是汉字信息输出的基本流程图。

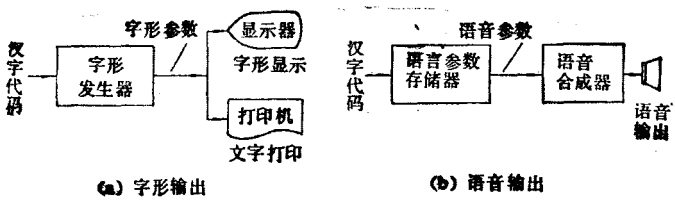


图 4·1　汉字信息输出基本流程图

图中（a）是字形输出，（b）是语音输出。语音输出技术在公众服务系统的自动查询和应答（如市内电话号码查询等）、自动检测和告警以及自然语言的机器翻译系统等场合，具有广泛的应用价值，但在汉字终端中尚较少应用。有兴趣的读者可参阅有关著作。

汉字字形的存储和输出技术，在汉字终端技术中占有重要的地位。能够存储和输出字形，这是任何一种文字终端设备都具备的功能，但是，汉字终端的字形存储和输出技术要比西文终端复杂得多，这是长期以来阻碍汉字信息处理进入机械化、自动化之门的主要原因。

科学技术的进展解决了汉字存储和输出技术中的许多难题。超大规模存储器芯片的研制和应用，新型显示器件和打印技术的开发，使汉字终端彻底改变了庞大而笨重的旧貌，迈进了与西文终端并驾齐驱的发展大道。汉字存储和输出技术的进步，对促进汉字信息处理、办公自动化、汉字通信等领域的发展具有很重要的意义。

本章先介绍存储器的基本知识以及字形发生器的有关内容，然后介绍汉字显示器和汉字打印机。

4·1 存储器概述

从计算机硬件角度来看，字形发生器的主体是一个存储字形信息的存储器。因此，有必要对各种存储器的性能、特点等基本知识作些概要介绍。

1. 存储器的分类及特点

从存储介质的不同类型分，微型计算机的存储器主要有半导体存储器和磁盘存储器两大类。光盘存储器近年发展很快，是一种很有应用前景的大容量信息存储器。

（1）半导体存储器

半导体存储器是一种用半导体材料以集成电路工艺制成的

存储数据信息的电子器件。半导体存储器的主要优点是：①存储单元阵列和主要外围逻辑电路制作在同一个硅片上，输入和输出电平可以同片外电路兼容和匹配，可使存储器与计算机其它电路单元的接口大为简化；②数据的存入和读取速度比磁盘存储器约快三个数量级，可大大提高计算机运算速度；③利用大容量半导体存储器可使存储体的体积和成本大大下降。

用作计算机主存储器的半导体存储器，是1970年前后开始生产的1K位动态随机存储器。随着大规模集成电路工艺技术的发展，现在几兆位存储容量的半导体存储器已经大批量生产，并应用于汉字字形发生器等存储数据信息量很大的领域。

按照制造工艺技术和材料的不同，半导体存储器有双极型存储器和MOS存储器两种。双极型存储器是以双极型晶体管为基本元件用双极集成工艺制成的，它有TTL（晶体管——晶体管逻辑电路）和ECL（发射极耦合逻辑电路）两种型式。由于双极型晶体管开关时间极短，故双极型存储器的数据存取速度比MOS存储器要快得多，适用于需要高速运算的场合。

MOS存储器即金属氧化物半导体存储器，它是一种以金属氧化物半导体场效应晶体管作为基本元件的存储器。按制造工艺的不同，有NMOS（N沟道金属—氧化物—半导体集成电路）和CMOS（互补金属—氧化物—半导体集成电路）等数种。MOS存储器的成本低，功耗小，可以做成极高的集成度，是汉字字形数据的理想存储体。

按功能不同，半导体存储器可分为三类：随机存储器、只读存储器和串行存储器。

① 随机存储器

随机存储器的全称是随机存取存储器，简称为RAM。它

的特点是对于存储器中任意一个地址单元的内容（数据），都可以在使用中随机地高速写入或读出。当写入数据时，原有的存储内容就自动消失；当读出数据时，存储内容仍保留不变。随机存储器的存储单元内部结构一般是组成二维方矩阵形式，即一位一个地址（如64K×1位）；但有时也有编排成便于多位输出的形式（如16K×4位，8K×8位等）。随机存储器主要用于组成计算机的主存储器等要求快速存取的场合；在汉字系统的混合型多级字形发生器中，也用来建立临时字库。

按工作方式不同，随机存储器又可分为静态和动态两类。静态随机存储器的单元电路是一个触发器。可规定A或B两个晶体管中的一个导通时，代表“1”或者“0”，作为存入的数据信息。只要有足够高的电源电压存在，存入的数据状态便不会改变，不需要进行刷新，因此使用比较方便。

动态随机存储器的单元电路是由一个MOS电容和一个MOS晶体管构成，数据以电荷形式存放在电容器之中，一般以无电荷代表“0”，有电荷代表“1”（反之亦可）。单元电路中的MOS晶体管是一个开关，它控制电容器中电荷的存入或取出。通常MOS电容与其相联接的PN结有微弱的漏电作用，因此动态随机存储器需要每隔2～4毫秒对单元电路存储的信息重写一次，称为刷新。动态随机存储器由于单元电路器件数量少，集成度高，应用最为广泛。

② 只读存储器

只读存储器是用来存储固定的数据或信息的存储器，简称为ROM。它的特点是在使用过程中只能读出信息，不能写入或更改。即使断开电源，信息也不会丢失，通电后可以立即使用。只读存储器的存储单元只有一个二极管或三极管，一般规定当管子接通时为“1”，断开时为“0”（反之亦可）。

按写入数据的方法不同，只读存储器分为以下三种类型：

掩膜式ROM，又称为MASK ROM或MROM。它是在工厂生产设计只读存储器掩膜版时，就将数据编写在掩膜版图形中，光刻时便转移到硅芯片上。因此出厂封装后用户只能读取已存入的数据，自己不能再重新编写。掩膜式ROM的集成度高，适合大批量生产，价格便宜。它作为国标汉字点阵数据的存储芯片，已经广泛应用于各种汉字设备中。

可编程ROM，简称为PROM。它是在出厂后由用户根据需要自行写入数据的只读存储器。而且，写入以后具有永久固定的内容，不能再重新改写，只能读出。一种常用的PROM是熔断型的，就是将串接在每个存储单元（如肖特基二极管）上的熔丝烧断的方法，来决定该单元的数据是“0”还是“1”。

可改写的ROM，简称为EPROM。它也是由用户根据需要写入数据的存储器。在使用过程中与其它两种ROM一样，只能读不能写。不同的是EPROM中的信息，还可以用专用设备“擦掉”（抹除）重写。所以在用户调试程序过程中，使用十分方便。许多汉字终端的专用型字形发生器，也常用EPROM作为汉字字形数据的存储器。

按照抹除信息的不同方式，EPROM有紫外线抹除，X射线抹除和电抹除等多种类型。一经抹除，存储器的全部单元数据为“1”。重新写入时，可以分多次写（甚至一次只写一个单元），也可以一次写入全部单元。一块芯片的抹除写入次数可达上百次。

③ 串行存储器

串行存储器的存储单元列成一维结构，数据按每列顺序读取，如移位寄存器和电荷耦合存储器等。

现将MOS存储器的分类归纳如图4·2所示。

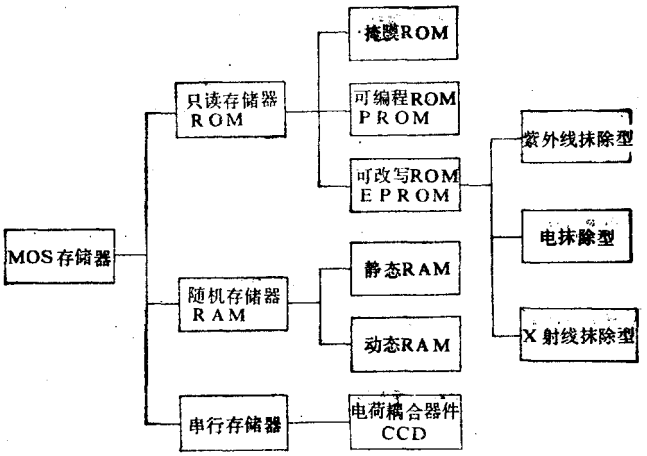


图 4·2 MOS存储器分类示意图

（2）磁存储器

磁存储器是利用物质的磁性来记录（存储）信息的器件。

早期的磁存储器是磁芯存储器。它是用具有矩形磁滞回线的铁氧体材料制成的微小型磁环，具有两个稳定的剩磁状态，可分别代表“1”和“0”。它的特点是不需要外加功率来维持剩磁状态，写入和读出都比较方便。由许多磁芯元件组成的磁芯存储器，曾作为内存储器在计算机应用初期发挥过重要作用。我国六十年代末研制的中文电报译码机也曾用它作为汉字字形数据存储器。但由于磁芯存储器很难进一步缩小体积，自七十年代开始逐步被半导体存储器所取代。

现代应用的磁存储器是一种磁表面存储器。它以涂覆于非磁性材料表面的磁性薄层作为存储介质，包括磁鼓、磁盘、磁

卡片和磁带等存储设备。由于磁表面存储器的存储容量大，记录（写入）和读出方便，已广泛应用于录音、录象、计算机外存储器等领域，在汉字终端和汉字微机系统中，主要使用磁盘和磁带存储器。

磁盘存储器是目前应用最广的磁存储器之一。它包括磁盘和驱动器两部分。磁盘是一种表面涂敷有磁记录介质的圆形薄片。工作时在驱动器的带动下高速旋转，通过在磁盘表面作径向运动的磁头，能在磁盘的任意位置上记录或读出信息。

磁带存储器是以磁带为存储介质的磁存储器，由磁带、传动机构及磁头等组成。磁带是将磁粉涂于塑料带基上面制成，当装入传动机构后，与磁头作相对运动。通过磁头的电磁转换作用，在磁带上顺序记录或读出信息。高档的磁带存储器价格昂贵，常作为中、大型计算机的外围设备之一。一种小型的盒式磁带录放机，由于价格低廉，使用方便，也可作为普及型汉字终端的辅助存储器。

现将两种磁存储器的性能特点比较列于表4·1。

表4·1　　两种磁存储器性能比较

类　别	磁盘存储器	磁带存储器
容　量	1亿字符/盘组	1千万字符/卷
存取时间	几十～一百毫秒	以秒计
存取方式	随机存取	顺序存取
价　格	较　贵	便　宜
使用环境	要求较高	一　般

除此以外，还有一些存储器件，如光盘和磁泡等，也可作为数据存储器。光盘的存储介质是表面上有一层极薄镀层的塑料薄膜。写入时激光把薄膜表面熔成凹坑，读出时用能量较小的激光检出信息孔。由于激光的聚集性好，可以获得很大的存

储容量，是未来大容量汉字库的理想存储体。

2. 半导体存储器的基本工作过程

半导体存储器（MOS存储器）是计算机的内存储器，也是汉字终端字形发生器的基本存储体。现以一个简单的半导体存储器为例，来说明存储和读取数据的基本过程。

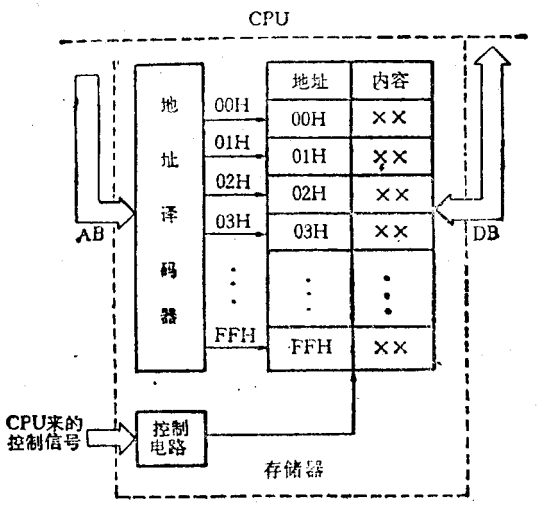


图 4·3 简单的半导体存储器结构图

图4·3所示一个简单的半导体存储器结构示意图。这是具有256个存储单元的存储器，存储单元的地址编号为00H～FFH（FFH是16进制数，等于10进制数255），每个存储单元存放8位二进制数，称为一个字节。存储单元的地址和这个地址中存放的数据，都可用二进制或16进制数来表示。图中“地址”和“内容”的含义是不同的，两者不能混淆。

半导体存储器通过地址总线(AB)、数据总线(DB)、控制总线(CB)与CPU相连。当需要读取或写入数据时，CPU首先通过地址总线、并经过地址译码器指明需访问的存储单元地址号；同时用控制总线经过控制电路指明是“读出”或“写入”；如果是“读出”，该地址单元存储的内容便出现在数据总线上，供CPU读取；如果是“写入”，CPU把出现在数据总线上的数据存放到该地址单元内。

对半导体存储器来讲，地址译码方法是正确实现数据存取的关键。图4·3所示的256个存储单元的简单存储器，只需8根地址线，用8中选1的译码器便可实现地址译码。但随着存储容量的增大，这种编址方法将使译码器的逻辑电路变得十分复杂，于是，就有图4·4所示的二维编址方法。由图可知，存

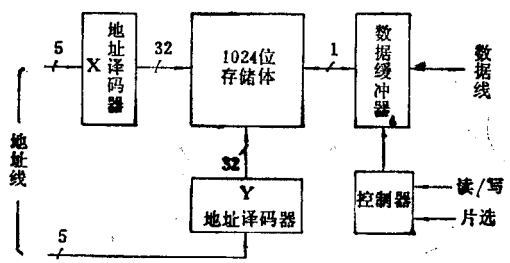


图 4·4　1024位存储器的编址方法

储器的10根地址线分成两组，X、Y地址译码器各接5根，可以选出X方向和Y方向各32线中的一条线。再从X、Y线相交点选出1024位存储单元中的一位。图4·5是用作字形发生器的1兆位芯片外观引线和内部结构图。

3．磁盘存储器的基本工作过程

磁盘存储器是利用磁记录技术，在旋转的圆盘上直接存取

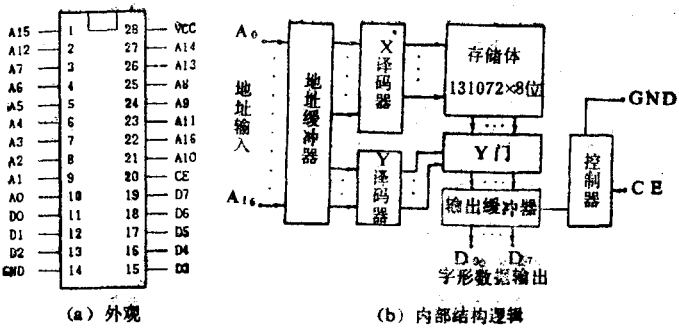


图 4·5　1兆位MASK ROM外观引线及内部逻辑图

数据的存储器，是目前广泛应用于计算机的外存储器，在汉字系统中也用作字形发生器的存储体。

磁盘存储器是由磁盘、驱动器和控制器三部分组成。

磁盘是表面涂着磁性材料的圆盘，靠磁性材料介质的磁化方向不同来存储信息。磁盘可按盘基的材料不同分为硬磁盘和软磁盘两类。汉字终端常用的是薄形的133毫米(5.25英寸) 和89毫米 (3.5英寸) 的软磁盘。图 4·6 是一种133毫米的软磁盘外形。磁盘本身被封装在一个方形保护套内。套上有三个小孔：

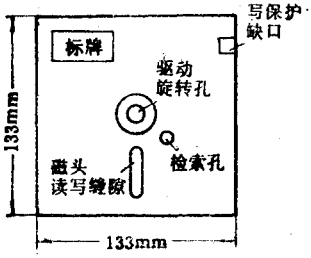
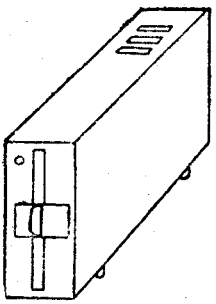


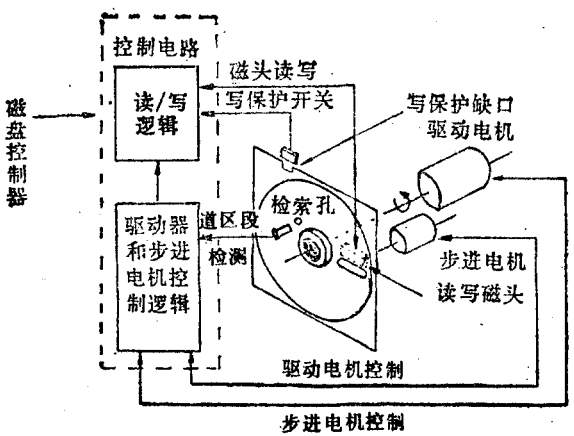
图 4·6　133毫米软磁盘外形图

中心大孔是驱动旋转用的，小孔是检索起始位置标记用的，长形孔是供磁头寻找读写磁道用的。右上方的缺口是为保护信息用的，若此缺口被封住，就不能对此盘再写入新的信息。

磁盘驱动器是驱动磁盘旋转并进行读写操作的装置。它的外形及内部结构如图4·7所示。



(a) 外形



(b) 内部结构

图 4·7 磁盘驱动器外形及内部结构图

磁盘驱动器主要由读写磁头、驱动电机、步进电机和控制电路等部分组成。驱动电机带动磁盘以每分钟300转左右的恒速旋转，步进电机则带动磁头沿磁盘的径向移动，在找到规定的盘面位置（区、段号）之后，控制电路根据磁盘控制器送来的命令，指示磁头进行读写数据的操作。

磁盘的存储容量比较大，一个133毫米软磁盘的存储量达320K～1.2M字节。温式硬磁盘的容量已达40兆字节。但磁盘存储器的存取速度较慢，在接通电源后，约需1～2秒，待驱动电机转速稳定之后方可进行读写操作。在读写过程中还要通过磁头的机械动作才能移到所需的磁道位置。在工作时，磁盘驱动器应处于洁净环境中，以防灰尘微粒侵入把磁头与盘面划伤。

4·2 汉字字形发生器

字形发生器又称字形库，是一种专为显示、打印设备提供字形数据的专用存储器，它可以根据输入代码输出相应的字形数据，是任何文字成形设备不可缺少的部件。对西文来说，由于西文字母数量只有26个，加上数字、符号和小写字母总数不过一百多个。而且字符的形状不复杂，因此，西文文字成形设备的字形发生器存储容量小，结构简单，在设备中不占重要地位。

汉字字形发生器的情况则不同。由于汉字数量大，字形结构复杂，各种字体变化多，致使汉字字形数据量极大，因而汉字字形发生器在原理、结构等方面与西文字形发生器相比都有很大的差别，在汉字成形设备中占有很重要的地位。如果没有汉字字形发生器，任何终端都不可能显示或打印出汉字。字形发生器质量如何，对汉字文本质量和汉字系统的效率有非常密切

的关系。

1. 字形发生器的分类及特点

按照存储信息的不同形式，字形发生器可分为模拟式和数字式两大类。

（1）模拟式字形发生器

模拟式字形发生器是将整个字形以模拟信息的形式记录在存储体上。读取时，可以通过机械的或化学的方法直接在纸面上印出字形；或者用光电转换和扫描的方法把字形变成电信号，再进一步显示和打印输出。一个简单的实例，是一种具有菊花瓣字盘的电子式打字机。它把需打印的字母、数字和符号直接铸压在菊花瓣状圆盘的顶端，这种菊花瓣字盘装在一个能左右移动的小车上，并且可以灵活地旋转。工作时，印字机构的控制电路通过对字形信息的分解，控制字盘的旋转角度及小车的左右移动位置，在适宜时间迅速击打字盘的顶端，便能逐字逐行地打字输出。

早期的汉字终端也采用过模拟式字形发生器来提供汉字字形。例如把几千个常用汉字字形刻铸在一个圆柱形的字鼓上，这个字鼓能够灵活转动，并能与打印头作相对的轴向移动。终端的控制电路把汉字信息分解成控制字鼓旋转和打印头位移的参数，当对准所选字形时，打印头迅速一击，便能直接打印出汉字。

以上讨论的字盘和字鼓，实质上都是一种记录了字形模拟信息的机械式的字形发生器。在西文终端中，由于字符数量少，字形也较简单，灵活轻便的菊花瓣式字盘打印机至今仍在广泛应用。但对于汉字终端来说，几千个汉字构成的庞大字鼓，使

设备体积大，速度慢，故很难推广使用。

以后，人们又设计研制了以光学字模为特点的多种模拟式字形发生器，并成功地获得了应用。这类字形发生器大体可分成两种：一种是用缩微照相的方法，把汉字及其它字符的字形记录在一个透明的圆盘上。工作时，用闪光灯照射旋转的文字盘，根据所需字形在盘上的位置控制发生闪光的时刻，便能得到该字形的影象，再经过透镜系统投影到印字机内的感光面上成象输出；或者以同样的方法投射到摄象管的光电面上，通过摄象管对字形逐点扫描，便能把字形图象变成电信号，再去控制显示器或打印机输出。这种方法的输出速度约为每秒二、三十个汉字，常用在需要高质量字形的普通印刷设备上。

另一种是用光刻方法将字形记录在固定的文字模板上。工作时用飞点扫描器的电子束扫描，从文字模板中选出所需字形，再叠加上微区域扫描，形成文字的光栅图形，经聚光镜输入光电倍增管，变成电信号后去控制显示器或打印机输出；或者用激光全息摄影方法将字形记录在全息光存储器中，工作时用激光偏转器的激光束扫描选字输出。这种方法不但能获得高质量的字形输出，而且输出速度高达每秒500～3000个汉字，常用在激光照排系统及高速激光印字输出设备上。

以上两种应用模拟式字形发生器的字形输出设备能输出与标准字体一样的高质量汉字，但是设备相当复杂，价格比较昂贵，而且字形的种类和字数一经确定就难以变更。随着数字式字形发生器技术的发展，输出字形质量的优化，以及成本的大幅度下降，模拟式字形发生器的使用范围也就日益缩小了。

（2）数字式字形发生器

数字式字形发生器是将字形分解成数字形式（数据）记录在

存储介质上的一种设备。按照数据的不同性质，数字式字形发生器可分为两种：一种是点阵式字形发生器，它存储的是汉字字形的全点阵数据；另一种是压缩式字形发生器，它存储的是汉字部件或字形的压缩数据。

由于把字形变成数字形式，因此数字式字形发生器可以不用复杂、昂贵的光学机电器件作为存储体，而代之以各种通用、简便的数据存储器作为存储介质。这一改变，不但使字形发生器的结构大为简化，成本大大降低。而且使用极其方便，设备可靠性大大提高。随着超大规模集成电路工艺技术的进步，数字式字形发生器已经得到了普遍应用。在本节的以下各小节中，将对数字式字形发生器的原理和应用作较详细的介绍。在下文中凡提到字形发生器时，未专门说明其类型的，都是指数字式字形发生器。

按照存储体的不同结构，字形发生器又可分为磁盘式、混合式、ROM式三种类型。各种类型的结构如图4·8所示。

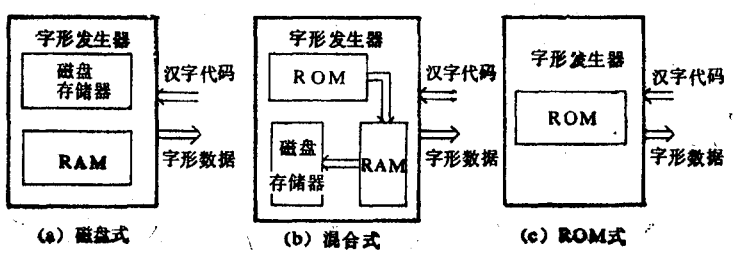


图 4·8 字形发生器的结构类型

① 磁盘式字形发生器

磁盘式字形发生器是把汉字字形数据分区存放在磁盘存储器中，其结构如图4·8（a）所示。由于磁盘的存储容量大，故

适用于需用汉字量很大的系统，作为共用的汉字库。但磁盘驱动器的存取速度较慢，使用磁盘式字形发生器时，需在系统内部开辟一个专用RAM存储区。在系统启动时，先将磁盘存储器中的常用字一次调入RAM区中，建立临时字库。如果临时字库能容纳三千个左右汉字字形数据，可保证使用过程中99%以上的字形数据在内存读出，只当遇到极少数生僻字才到磁盘存储器中查找。这样可避免频繁地启动磁盘驱动器，提高字形数据的读出速度。

② 混合式字形发生器

混合式字形发生器是把数量较少的常用汉字的字形数据存放在ROM中，大量的次常用字和罕用字存放在磁盘存储器中，其结构如图4·8（b）所示。混合式字形发生器工作时，调用磁盘存储器的方式有两种：一种是开机后立即启动驱动器，把次常用字调入内存的专用RAM存储区，建立临时字库；另一种是不建立临时字库，遇到ROM中查找不到的字时再启动磁盘驱动器查找。前一种方法的优点是可以减少读取磁盘数据的次数，输出速度快；而且ROM的容量也可以小一些。但是，每次开机后要进行一次数据装填操作，并占用较大的系统内存。后一种方法不占用系统内存，但ROM的数据存储量要大一些，而且读取磁盘存储器数据的次数较多，输出速度较慢。

③ ROM式字形发生器

ROM式字形发生器是把国标基本集（GB2312—80）中的七千多个汉字和非汉字字符的字形数据全部固化（存放）在MASK ROM或EPROM中，故又称为固化字库，其结构最简单，如图4·8（c）所示。工作时，系统根据地址码或国标码直接访问ROM存储器，字形数据输出速度快；不需装填操作过程，不占系统内存，断电时数据不会丢失。因此ROM式字

形发生器的使用最为方便。与上述两种字形发生器相比，ROM字形发生器由于存储器容量有限，故收字数量较少，一般只限于国标基本集的汉字以及少量的增补汉字和字符，更多的字暂不能收入。这对于国内通用型的汉字终端和汉字计算系统来讲，已能基本满足使用需要，对极少量专业用字和罕用字，则可通过增加系统的造字功能来解决。

由于超大规模集成电路工艺技术的新成就，以及国家标准点阵字模集的制订和颁布施行，现在单个芯片存储容量达4兆位（512K字节）的ROM式字形发生器已经大量投放市场。毋庸置疑，ROM式字形发生器的应用范围必将进一步扩大，并且对汉字终端的普及推广产生意义深远的影响。

2. 数字式字形发生器的原理

如前所述，数字式字形发生器是用数据存储器来存放字形信息的，那末，怎样把汉字字形变成数据？数据在存储器内是以怎样的格式存放的？怎样实现字形数据的压缩存储等问题，将是本小节讨论的主要内容。

（1）汉字字形的数据表示及存储

汉字字形是一种连续变化的模拟量，为了能够用二进制的“0”、“1”数据来表示，就需把笔画组成的字形转换成由点子组成的点阵字形。

怎样把笔画字形变成点阵字形？我们可以做一个简单的实验。用黑纸刻画出一个汉字或字母，把它置于有网状栅格的纸上。要求整个字形基本布满一个小方块，小方块内的网状栅格的密度与字形笔画的粗细大体可比。仔细分辨字形笔画与每一个栅格的关系：对于全部或大部分被笔画复盖住的栅格，用一

个黑点表示；对于没有或基本上没有被笔画复盖的栅格，用一个空格表示；对于难以区别的“模棱两可”的小栅格，则根据整个字形的笔画特点来决定用黑点还是用空格表示。这样，就在字形所在的小方块平面内得到一个以黑点子描出的近似字形，这就是汉字的点阵字形。

显然，点阵字形与原笔画字形的近似程度，取决于字形的大小和字形方块内网状小栅格的数目。对于字形方块面积在5×5 mm^2以内的一般印刷字体，需用16×16的栅格网才能大体上正确描绘出常用汉字的点阵字形。用24×24的栅格网描绘的点阵字形，能够体现出字形的某些笔画顿点和笔锋，其相似程度已比较满意了。对于非汉字的字母、数字及常用符号，由于字形比较简单，一般只需5×7或者7×9的点阵字形就能较好体现原来字形。汉字16×16点阵字形和24×24点阵字形以及字母7×9点阵字形如图4·9所示。

在画出了汉字的点阵字形以后，便可以方便地得到它的点阵字形数据。我们规定以二进制值“1”表示黑点，“0”表示空格。从字形的左上角开始，按自左至右、自上而下的次序取数，就得到字形点阵的二进制数据值，数据的每一位直接对应每一个点。为了书写方便，通常用16进制数来表示字形数据，一位16进制数代表4位二进制数。于是可得到“啊”字（16×16点阵字形）、“熬”字（24×24点阵字形）和“A”字母（7×9点阵字形）的字形数据值如图4·9中所示（标在每一个字形左测）。“A”的字形数据为00 08 14 22 22 3E 22 22 00；“啊”的字形数据为0000 0F7E E904 A904 AA04 AA74 AC54 AA54 AA54 A954 E974 AD54 0A04 0804 0814 0808。“熬”的字形数据有144位16进制值，已在图中标明，不再列举。

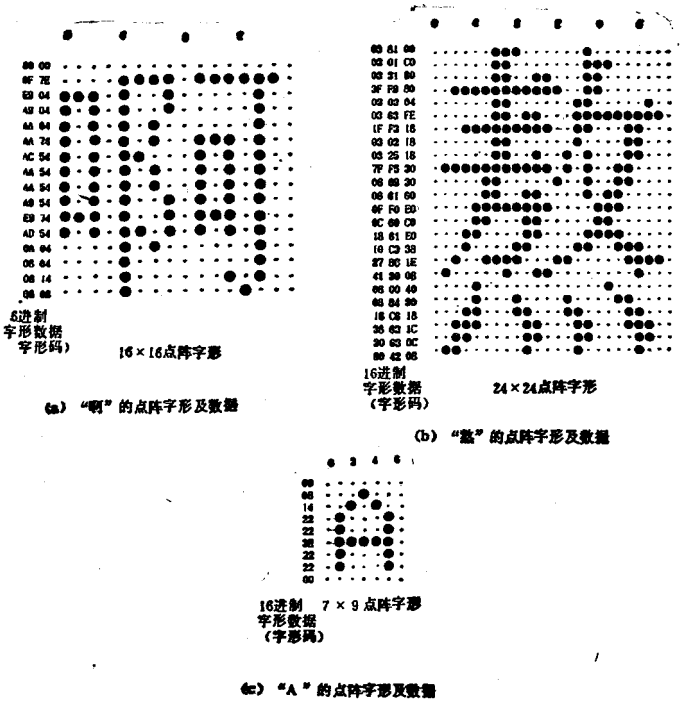


图 4·9 汉字及字母的点阵字形及数据

按照上述关于点阵字形设计的基本原理，已制订了对国标基本集15×16、24×24和32×32三种标准点阵字形及其字形数据，并分别以国家标准GB5199、GB5007和GB6345颁布实施。这是一项十分浩繁的基础工程。它对于汉字信息的存储、交换、输出都具有重要的作用。

有了字形数据，汉字的字形存储问题就可用通用的数字存储器来解决。通常半导体存储器是以字节为单位来存放数据

的，一字节为 8 位二进制数，即 2 位16进制数，因此，对于16×16点阵字形来说，一个汉字的存储量为32字节，24×24点阵字形的存储量为72字节，32×32点阵字形的存储量是128字节。总之，汉字的字形点阵数目愈大，存储一个汉字所需的存储器容量就愈大。

把一个汉字字形占的方块面积中每一个小栅格，对应存储器中的每一位存放的字形发生器，称为点阵式字形发生器。点阵式字形发生器的原理比较简单，只要规定好有笔画的黑点为“1”，没笔画的空白点为“0”，一点占一位，存入ROM即可。读取时也很方便，按位顺序读出数据后，送入显示设备的辉亮控制电路或点阵式打印设备的打印针驱动电路，就能把点阵字形显现出来，不需其它复杂的变换和处理过程。因此在大容量存储芯片价格不断下降的今天，点阵式字形发生器得到了广泛应用。

但是，点阵式字形发生器有个明显的缺点，就是每个汉字所需的存储量比较大。其存储量随字形点阵数目增加而增大，也就是说，要求字形质量愈高，每个汉字的存储量愈大；而且汉字的数量多，就以国标基本集的七千多汉字和图形字符计算，用24×24点阵字形，字形发生器的存储量高达500K字节(约4兆位)。在正式出版的报纸刊物中，还需要其它各种各样的字体和更高的字形质量，则整个字形发生器的存储容量可能要高达几百甚至几千兆位！这样巨大的数据存储量对于任何一个实用的微机系统都是无法忍受的，若使用大容量的硬磁盘系统来做字库，则字形输出速度将下降到每秒一、二十个汉字以下，无法实际应用。因此，在正式出版物的印刷系统中，字形数据的压缩是十分必要的。

从另一方面讲，汉字笔画的繁简程度差别极大，不论什么

字，一律用几百个、甚至几千个点子去描述，信息的冗余量是相当大的。实践证明对于大多笔画比较简单的字，在指定的方块里，有笔画的地方总要比空白的地方少得多。也就是在读取点阵数据时会出现大量连续为“0”的情况。这种信息冗余量大量存在的现象，给字形数据压缩提供了可能性。

（2）汉字字形数据的压缩

如前所述，汉字的字形结构有单字、部件、笔画三个层次，即由笔画组成部件，由部件组成单字。显然，组成单字的笔画或部件的数目，要比“点子”的数目少得多。由此可以联想到，用笔画或部件的特征数据去描述各种不同的汉字字形，比用点阵数据描述所需的数据量要小得多。于是就有了两种常见的字形数据压缩方法：笔画向量法和部件组合法。

①笔画向量法

笔画向量法是一种向量式字形表示法。它的基本原理是把一个汉字字形分解成一系列直线段，每个直线段用相应的一组向量数据表示。对于弯度比较大的“撇、捺、折”等笔画，则应分成若干直线段来近似地表示。具体的表示方法是：规定从某一直线段的始点到终点为一组向量数据；带弯、勾、折的笔画，应分成几组向量数据；对两个不连续的笔画，从上一笔画的终点到下一笔画的始点也作为一组向量数据。但这组数据应加标志，注明它是一根不画出来的“虚线段”。现以“汉”字为例，说明笔画向量的字形表示方法，如图4·10所示。

由图可见，设“汉”字所在平面是一个直角坐标系，左上角O点是原点。在X、Y方向各取32等分的小单元（即量化范围为1～32），用从原点开始的14个向量线段oa、ab……mn来描述。规定每一个向量线段的始点和终点的坐标值取量

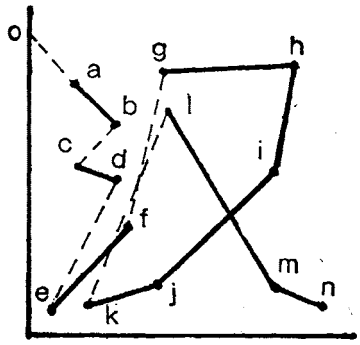


图 4·10 “汉”字的向量表示

化单元的整倍数。这样，可以得到在X、Y两个方向上的增量Δx和Δy为：

$$\Delta x = x_i - x_{i-1}$$

$$\Delta y = y_i - y_{i-1}$$

式中 i 是 1 ～32之间的正整数。按照书写习惯，规定笔画方向从上到下或从左到右时，Δx或Δy取正值，笔画方向从下到上或从右到左时，Δx或Δy取负值。则可计算出图4·10所示“汉”字的Δx和Δy值如下所列：

线段	ΔX	ΔY	虚实线段标志
oa	4	4	0
ad	4	4	1
bc	-5	5	0
cd	4	2	1
⋮	⋮	⋮	⋮
mn	5	2	1

虚实线段标志为“0”时，表示该线段是“虚线段”，不

画出来；为“1”时则为实线段。根据上述规定，Δx和Δy是带符号的0～31之间的整数，就可以用6位二进制数表示。加上标志位一个向量线段只需16位二进制数（即2个字节）已能代表。如图4·11所示。图中12～15的标志位，除了作为

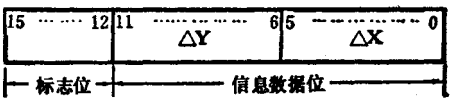


图 4·11 一个向量线段的数据表示

“虚实线段标志”的1位外，余下几位还可以作为向量表示法的扩展功能的标志位。例如，把某些最常用的复笔笔画和部件，如“口”“木”“氵”等，定义成专用的双字节笔画码，则使用起来更为方便。

笔画向量法的数据压缩量较大。以一个汉字字形平均用14个向量线段计算，只需2×14=28个字节数据就可表示，比24×24点阵字形的数据量压缩约2.5倍。在把常用部件定义为双字节笔画码后，字形发生器的数据压缩比还可进一步提高。

② 部件组合法

部件组合法又称叠加法，是实现字形信息压缩的另一种常用方法。

如前所述，在数量众多的汉字中，除了少数独体字形以外，绝大多数是由部件构成的合体字形。据统计，几千合体字的构字部件大致有二、三百个，其中构字能力强的常用部件不过几十个。于是，就可以用存储构字部件及其结构数据的办法，实现字形信息的压缩存储。需要输出字形时，根据相应的汉字代码读出需用的部件及构字数据，再由软件生成字形数据。按照这一原理设计的部件组合法（叠加法）字形发生器，由于可以不必重复存储大量相同的部件数据，因而大大压缩了字形发生器的

存储量。而且，存储的汉字字数愈多，压缩比就愈高。

部件组合式字形发生器存储的汉字字形数据分为两个部分：一部分是部件信息数据。就是表示各种构字部件笔画形状的点阵数据；另一部分是构字信息数据，就是指明字形结构类型以及部件在字形结构中所处位置的信息。例如，在部件组合式字形发生器的存储单元中，“杏”字所在单元存放的数据是：①“口”和“木”两个构字部件的点阵数据，（通常是存储这两个部件点阵数据的地址）；②指明该字是上下型，且“口”部件在下的信息数据。于是，当系统向字形发生器输入“杏”字的内部码，读出上述两组信息数据后，便可通过软件生成“杏”字的点阵字形数据输出。根据同样的原理，在“呆”字“困”字的存储单元存放的数据，第一部分是相同的，第二部分则不同。“呆”的构字信息数据应指明“上下型，‘木’在下”；“困”的构字信息数据应指明“内外型”。这样，在向字形发生器送入汉字代码后，便能有相应的字形点阵数据输出，不会搞错。

部件组合式字形发生器存放的部件信息数据与构字信息数据是相互有联系的。也就是说，同一部件在不同的字形结构位置中，往往具有不同的大小和形状。例如“口”部件在内外型字形结构中处于外围位置时，应该是大方“口”，在其它字形结构中就应是小“口”，具有完全不同的点阵数据。又如一个“木”，在作为单字或“林”，“森”组合字的构字部件时，尺寸应该有差别，点阵数据也是不同的。如果一种部件只有一个尺寸，势必在许多情况下无法构字，或者勉强构成字形，也是十分生硬呆板的。

由于汉字字形结构有较强的规律性，许多部件的构字使用频度很高，因此部件组合式字形发生器与笔画向量法相比有更

高的信息压缩比。国内生产的几种型号部件组合字库，包括国标基本集中全部汉字及图形字符的总存储容量为24K～32K字节，与点阵式字形发生器相比，存储信息数据量压缩8～10倍。而且，部件组合式字库的存储字数愈多，存储信息数据压缩比就愈高。

上述两种压缩式字形发生器的共同特点是字形数据的压缩和复原算法比较简单，硬件成本也比较低，适合于简易型的汉字终端设备使用。但是，共同的缺点是输出字形质量不好。主要原因是：用笔画向量表示字形时，对婉转自如的生动笔形用若干个小段直线去描述，总是显得生硬呆板；用部件组字时，各个部件之间的配合比例很难完全得当，也不容易做到部件尺寸大小完全适应不同字形的要求。因此，当超大规模集成电路的国标汉字点阵芯片大量生产、并且价格不断下降以后，上述简易型的压缩字库已日渐失去其应用价值了。

如前所述，字形发生器的字形信息数据压缩技术在排印书籍、报刊等正式出版物的精密照相排版系统中，仍然具有很大的实用意义。因为正式出版物的字体大小变化多，对字形质量要求很高。例如正文的5号字体要用108×108点阵字形，标题的1号字体要用576×576、甚至更大的点阵字形。而且，还要有宋体、仿宋体、黑体、楷体以及这些字体的变形共十多种，每种都需几千个字。存储以上字体字型的点阵式字形发生器，如果不用压缩技术加以处理，其存储容量之大是任何系统都无法承受的。

但是，保证高质量字形输出的精密字库信息压缩技术，是比较复杂的课题，国内外都投入了很大力量进行研究。近几年来，国内不少单位的研究取得了很大进展，并已进入实用阶段，例如，北京大学采用轮廓折线方法的字形压缩、放大技术是成

功的。他们在字形发生器总容量不到5兆字节的存储空间内，可产生多种不同字体的高质量字形，并已成功地应用在精密激光照排系统中，印出的字形优良，字号范围广，由于采用专用解码复形电路，实现了高速汉字输出。此外，“高次样条函数法”、“圆锥曲线近似法”、“圆弧曲线近似法”等字形数据压缩算法，分别以高次样条函数、圆锥曲线或圆弧曲线去拟合、逼近汉字笔画形状的轮廓线，从而能产生字形质量优美、逼真度很高的汉字输出。中南工业大学用“差分法”描述汉字笔形轮廓线，做到在不足500K位的存储空间内产生多种字体和字号的汉字，并在IBMPC微机上用软件解码还原字形，可以获得实用的汉字输出速度。还有一种“混合型”汉字信息压缩和还原技术，也都达到了一定的应用水平。

3. 数字式字形发生器的应用

如前所述，汉字字形发生器是汉字终端的重要组成部分。对它的正确应用是关系汉字终端设备性能及效率的重要问题。

（1）数字式字形发生器的主要性能指标

汉字终端及通用型的汉字微机系统已经都采用数字式字形发生器，其中绝大部分又是点阵式字形发生器。为了便于应用，首先讨论数字式字形发生器的字数及存储容量、字形质量、输出速度等几项主要性能指标。

① 字数及存储容量

一台汉字终端设备总共能显示或打印出多少汉字，取决于字形发生器存储了多少个汉字字形。因此，设计或采购设备时应该根据使用要求恰当地选择。

对于通用型的汉字设备来讲，由于大多已采用国标汉字点

阵芯片作为字形发生器，故能够提供国标基本集GB2312中的汉字6763个，非汉字图形字符682个。此外，根据4M位芯片存储容量作了一些增补和调整。例如，GB5007A型24×24国标汉字点阵芯片，除了存储GB2312中全部内容外，还增补了以下内容：

- 18×24点阵的半角ASCII码字符94个；
- 15×16点阵的汉字纵书符号16个；
- 各种控制符35个；
- 制表符15个；
- 汉字122个；
- 小写罗马数字符10个；
- 外文字符32个；
- 其它各种符号55个。

现将增补的122个汉字，32个外文字符、35个各种控制符、16个纵书符号和15个制表符，以及它们的国标区位码，列于表4·2、表4·3和表4·4中。

对于国标基本集和表4·2中没有收入的汉字，在需用时，则采取自造字形的办法解决。通过造字软件描画出该字的点阵字形，再将其数据存入指定的RAM区，作为用户自造字库取用。

对某些特殊专业用的汉字终端设备，也可在保证使用要求的前提下，采用只包括国标基本集的一级字3755个汉字及若干图形字符的固化字库；或者自行设计字数更少些的专用型EPROM（或PROM）简易字库，以力求减小设备体积和降低成本费用。

字形发生器的存储容量，取决于所存储的汉字数量和字形质量。存储汉字的数量愈多，所需的存储容量愈大，当存储汉

表4·2　　　　**增补的122个汉字表**

区位码	汉字	区位码	汉字	区位码	汉字	区位码	汉字
0664	甪	0692	关	1142	浲	1170	礽
0665	屃	0693	芗	1143	溠	1171	祎
0666	鄃	0694	奭	1144	溇	1172	祊
0667	伋	0878	圊	1145	滆	1173	祐
0668	佺	0879	岝	1146	溦	1174	禤
0669	佫	0880	峂	1147	漷	1175	牂
0670	侪	0881	峃	1148	潘	1176	研
0671	契	0882	峔	1149	澒	1177	硊
0672	冮	0883	峧	1150	澋	1178	碶
0673	邦	0884	崀	1151	澛	1179	磡
0674	邸	0885	崌	1152	澈	1180	磜
0675	郚	0886	嵖	1153	濛	1181	替
0676	郤	0887	嵡	1154	澴	1182	鸤
0677	鄀	0888	嶓	1155	瀞	1183	鹍
0678	鄚	0889	巂	1156	瀌	1184	鹐
0679	厘	0890	猺	1157	瀼	1185	鲂
0680	垞	0891	馇	1158	瀫	1186	雷
0681	垟	0892	汍	1159	弶	1187	鲊
0682	埲	0893	汈	1160	嫄	1188	鲖
0683	埈	0894	洸	1161	嫪	1189	鲞
0684	堌	1134	浉	1162	瑄	1190	鲇
0685	塆	1135	洺	1163	杧	1191	鲹
0686	瑕	1136	浐	1164	栟	1192	鳁
0687	堤	1137	洨	1165	栭	1193	鳛
0688	堋	1138	浘	1166	椑	1194	鞲
0689	窦	1139	涢	1167	殑		
0690	萩	1140	浛	1168	炯		
0691	弇	1141	涌	1169	炝		

国标码	内容
2B00H	犇
2B01H	吋
2B02H	舘
2B03H	堃
2B04H	浬
2B05H	瞟
2B06H	冇
2B07H	猕
2B08H	瑪
2B09H	舖
2B0AH	遊
2B0BH	祐
2B0CH	瓩

表4·3 增补的32个外文字符表

国标码	内容	国标码	内容
2800H	Ǒ	2810H	è
2801H	É	2811H	ò
2802H	Ä	2812H	ù
2803H	Ö	2813H	ì
2804H	Ā	2814H	ç
2805H	Ü	2815H	°
2806H	é	2816H	
2807H	ā	2817H	ī
2808H	ö	2818H	Ñ
2809H	ǎ	2819H	¿
280AH	ü	281AH	ñ
280BH	§	281BH	Æ
280CH	ß	281CH	Ø
280DH	£	281DH	æ
280EH	¥	281EH	ø
280FH	à	281FH	Ü

字总数确定后，输出的字形质量愈好，字形发生器所需的存储容量就愈大。表4·5列出了7445个汉字及字符图形的15×16、24×24、32×32三种点阵字形的字库存储量，以及需用芯片的数量。

② 字形质量

输出字形质量高低，是衡量字形发生器性能的一个重要指标，人们常常以打印文件的字形质量作为选用一台汉字设备的

表4·4　增补的35个控制符、16个纵书符号和15个制表符

国标码	内容	国标码	内容
2900H	^@	2910H	^P
2901H	^A	2911H	^Q
2902H	^B	2912H	^R
2903H	^C	2913H	^S
2904H	^D	2914H	^T
2905H	^E	2915H	^U
2906H	^F	2916H	^V
2907H	^G	2917H	^W
2908H	^H	2918H	^X
2909H	^I	2919H	^Y
290AH	^J	291AH	^Z
290BH	^K	291BH	^[
290CH	^L	291CH	^\
290DH	^M	291DH	^]
290EH	^N	291EH	^^
290FH	^O	291FH	^_
2921H	■	2923H	◢
2922H	◀		

(a)

国标码	区位码	内容	国标码	区位码	内容
2970H	0980	ㄱ	2978H	0988	ㄱ
2971H	0981	ㄱ	2979H	0989	ㄴ
2972H	0982	ㄴ	297AH	0990	ㄴ
2973H	0983	ㄷ	297BH	0991	ㄴ
2974H	0984	ㄷ	297CH	0992	ㄴ
2975H	0985	ㄹ	297DH	0993	ㄴ
2976H	0986	ㄴ	297EH	0994	ㅣ
2977H	0987	ㄴ			

(b)

区位码	内容	区位码	内容
0201	，	0209	ㄱ
0202	、	0210	ㄴ
0203	。	0211	ㄱ
0204	：	0212	ㄴ
0205	；	0213	ㄴ
0206	！	0214	ㄴ
0207	？	0215	ㄴ
0208	…	0216	ㄴ

(c)

表4·5　　三种点阵字形的字库容量及芯片数量

字形点阵数	单字存储量（字　节）	字库总存储量（字　节）	需用芯片数量（片）		
			1M位	2M位	4M位
15×16	32	238K		1	
24×24	72	536K	1*		1
32×32	128	953K			2

* GB5007C型国标汉字点阵芯片，可以只用1片4M位芯片，放下国标基本集的一、二级汉字及主要的字符图形（省略了该集的第4、5、6、7、8五个区的非常用字符图形）。

标准，因此选用高质量字形输出的字形发生器是一个很重要的问题。

如前所述，用笔画向量法和部件组合法的压缩字库，由于输出字型质量较差，故已很少使用；就点阵式字形发生器来讲，输出字形质量首先与点阵字形的设计有关，良好的点阵设计能使字形结构端正，笔画匀称合理，提高点阵字形质量。但是，从根本上讲，字形发生器输出的字形质量取决于采用哪一种点阵字形。目前在汉字终端及通用的汉字微机系统中采用的汉字点阵字形，不外乎15×16、24×24和32×32三种。

（A）15×16点阵字形

15×16点阵字形是在16×16网状栅格中描画的，可容纳汉字笔画的最大限度横笔是8笔，竖笔也是8笔。由于汉字的结构特性，不可能每个字形都在栅格方块中“顶天立地”。既要照顾到笔锋的舒展，又要尽量避免与上下左右笔画并线或粘连。故实际可用的笔画限度，横、竖都只有7笔。超过了上述笔画限度，就难以用15×16点阵表示原来的字形。

根据上述情况，国家标准点阵字形研制组在制定15×16点阵字形标准时，对横、竖超过8笔的138个繁笔字，参照中国传

统书法草书对繁笔字处理的方法，压缩变通笔形，进行了简化。设计时主要依据了两条原则：

·被压缩变通简化处理的繁笔字，保留了原字字形的特点；

·简化后的字形，看起来仍然逼真，不会被误认为是其它字。

众所周知，宋体字的最大特点是横细竖粗，笔画规范，字形端正，常用于印刷正式出版物。但表现粗画需要双画线，用15×16点阵是无法实现的，因此只能用横竖笔画都一样来设计字形。由于在设计15×16点阵标准时，按照上述原则在技术上采取了一定的措施，注意了笔形、字体的标准化和字形结构的规范化，所以设计的国标15×16点阵字形仍然保持了汉字的宋体结构和宋体风格。这种字体，目前已广泛应用于汉字显示终端及对字形质量要求不高的简易型汉字打印终端上。

现将国家标准15×16点阵字形的样张列于表4·6。

(B) 24×24点阵字形

在汉字字形笔画结构中，一般规律是横笔多、竖笔少。根据已公布的通用汉字印刷体规范字形，在国标基本集的一、二级汉字字形中，只有三个字的横笔笔画最多，为12笔。因此，用24×24点阵字形，可以全部做到笔画齐全、横细竖粗的双线体字形。其中绝大多数字形可以具有宋体字的笔画规范，结构端正、字形美观，均能满足办公自动化及情报资料检索等场合对字形质量的要求。随着超大规模集成电路和新型打印设备工艺技术的不断提高，24×24点阵字形已经越来越多地应用在各种汉字终端设备、中西文电子打字机以及汉字信息处理系统中。

现将国家标准24×24点阵字形样张列于表4·7。

(C) 32×32点阵字形

Retro Workshop
果粉工作室

表4·6　　国标15×16点阵字形样张

24区	3821	3822	3823	3824	3825	3826	3827	3828	3829	382A	382B	382C	382D	382E	382F	3830	3831	3832	3833
01/19		浮	涪	福	袱	弗	甫	抚	辅	俯	釜	斧	脯	腑	府	腐	赴	副	覆	赋
20/39	复	傅	付	阜	父	腹	负	富	讣	附	妇	缚	咐	噶	嘎	该	改	概	钙	盖
40/59	溉	干	甘	杆	柑	竿	肝	赶	感	秆	敢	赣	冈	刚	钢	缸	肛	纲	岗	港
60/79	杠	篙	皋	高	膏	羔	糕	搞	镐	稿	告	哥	歌	搁	戈	鸽	胳	疙	割	革
80/94	葛	格	蛤	阁	隔	铬	个	各	给	根	跟	耕	更	庚	羹					

25区	3921	3922	3923	3924	3925	3926	3927	3928	3929	392A	392B	392C	392D	392E	392F	3930	3931	3932	3933
01/19		埂	耿	梗	工	攻	功	恭	龚	供	躬	公	宫	弓	巩	汞	拱	贡	共	钩
20/39	勾	沟	苟	狗	垢	构	购	够	辜	菇	咕	箍	估	沽	孤	姑	鼓	古	蛊	骨
40/59	谷	股	故	顾	固	雇	刮	瓜	剐	寡	挂	褂	乖	拐	怪	棺	关	官	冠	观
60/79	管	馆	罐	惯	灌	贯	光	广	逛	瑰	规	圭	硅	归	龟	闺	轨	鬼	诡	癸
80/94	桂	柜	跪	贵	刽	辊	滚	棍	锅	郭	国	果	裹	过	哈					

表4·7　　国标24×24点阵字形样张

二．主要特点：

1．本数据集共提供汉字图形字符7451个。除138个繁笔字作简化处理外，数据的字形均符合我国文化部和中国文字改革委员会1964年颁布的《印刷通用汉字字形表》和国务院批准颁发的"汉字简化方案"，也符合中国国家标准GB2321－80《信息交换用汉字编码字符集基本集》中的汉字数量和排序。

2．汉字图形字符所用的点阵为15×16，第十六列全为空格，作为字与字间的间隔。每个汉字图形字符的点阵数据为16×16，计32个字节。

3．由于15×16点阵汉字字形受到点阵栅格少的限制，对纵横向超过八笔的138个繁笔字进行了必要的字形简化和规范化处理。

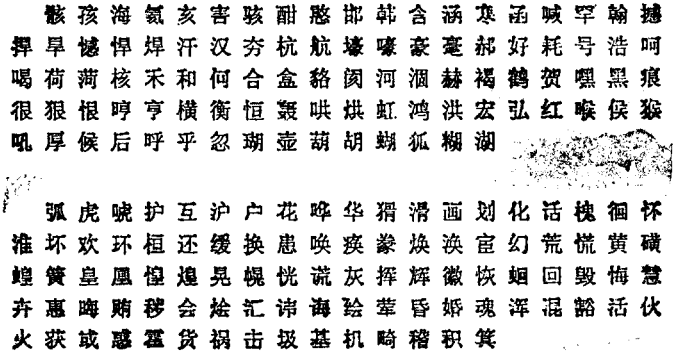
4．字形的大小一致。均匀足格，重心平稳，字距疏密适宜

5．整套字形体现出宋体字的风格。显示的字形美观。

32×32点阵字形的点数较多，可以正确表示包括繁体字在内的所有汉字字形；在通用汉字范围内，还可以做成多线体字形，比24×24点阵字形能更好地体现出宋体字风格，完整地表达出它的笔画特点。可以带有一定程度的笔锋，做到撇、捺自然婉转，舒畅流利。适用于高精度的打印设备（如台式激光打

印机）和高分辨率显示器。在24针和48针的针式打印机上也应用较多。现将国家标准32×32点阵字形样张列于表4·8。

表4·8　　国标32×32点阵字形样张



上述三种点阵字形的国家标准是《信息交换用汉字（15×16、24×24、32×32）点阵字模集及数据集》，即GB5199、GB5007和GB6345。

这一套标准点阵字形，是以宋体字为基本体，采用了微长形设计。字形标准、结构端正、笔画规范，部首归类相对统一，大小匀称，组合比例协调。它标志着我国汉字点阵设计达到了很高的水平。

⑧ 输出速度

字形发生器的输出速度，是指从输入一个汉字代码（或地址码）到输出该字的字形点阵数据所需的时间。

如前所述，字形发生器的输出数据主要供给显示和打印设备。对于打印设备来讲，由于总有一定的机械运动结构，对字形发生器的输出速度要求不高。目前，汉字打印机的打印速度很少超过每秒100个汉字，因此，字形发生器的读出速度只要

在毫秒级以上，就可以满足要求。对于显示设备的情况就不同了，它要求字形发生器有较高的数据读出速度，才能满足显示刷新的需要。

假设一显示器每屏显示480个汉字，为了保持显示屏字形的稳定，要求不停地向显示器提供相同的数据信号，称为屏幕刷新。如果显示器的屏幕刷新频率为50Hz，就是每秒读取屏显汉字的点阵字形数据50次，则字形发生器的输出速度为：

$$480 \times 50 = 24000 \text{汉字/秒}$$

已知24×24点阵字形的数据量为72字节/汉字，则可计算出字形发生器输出一个字节的时间应小于：

$$\frac{1}{24000 \times 72} = 5 \times 10^{-7} \text{秒} = 500 \text{毫微秒}$$

字形发生器的输出速度，首先取决于存储介质的不同类型。如前所述，半导体存储器的读取数据速度快（约200ns），因此，以半导体存储器为介质的数字式字形发生器，一般都能满足直接输出数据供给显示屏刷新的要求。当使用以磁盘存储器为介质的字形发生器时，则应事先把字形点阵数据读出，置于专用的显示刷新存储区中，以保证满足屏幕刷新的速度要求。

字形发生器的输出速度，还与点阵数据在存储芯片中的存放格式、存储器的软硬件设计等因素有关。一般说来，点阵式字形发生器的输出速度高于压缩式字形发生器，因为后者在输出时，还要通过一定的运算变换，把存储的笔画向量或部件参数复原成字形点阵数据。有关字形发生器存储体的数据存放格式及硬件线路连接等内容，将在下面字形发生器设计知识中作进一步介绍。

（2）字形发生器设计知识

应用半导体存储器芯片做成汉字字形发生器（即固化字库），是目前汉字终端及汉字微机系统最常用的方法。

在国标汉字点阵芯片大批量投产以前，人们通常用大容量的EPROM存储器芯片来自行设计固化字库。其基本的设计步骤是：

① 根据需用的汉字数量和字形的点阵数，确定EPROM芯片的容量及芯片数

假设字形发生器需存储的数量为：汉字H个，全角字符C_1个，半角字符$C_{1/2}$个，四分之一角字符$C_{1/4}$个*。字形点阵数为m×n，则所需半导体存储器容量为：

$$M=\left(H+C_1+\frac{1}{2}C_{1/2}+\frac{1}{4}C_{1/4}\right)\times m\times n \qquad (1)$$

式中，M的单位是“位”（bit）。应该注意对15×16点阵字形来说，点阵数通常取m=n=16计算。

按上式算出所需存储容量后，便可根据芯片的集成度C计算出所需芯片数S：

$$S=\frac{M}{C} \qquad (2)$$

式中，C的单位应与M相同，也取“位”（bit），S应对上取整数值，即当K＜S≤K+1时，应取S=K+1。

例如，设计一个普及型的固化字库，收集常用汉字3755个，半角ASCII码字符94个，字形点阵15×16，则根据（1）

---

* 对于m×n点阵字形来说，全角字符的字形点阵数为m×n，半角字符的字形点阵数为$\frac{1}{2}m\times n$。四分之一角字符的字形点阵数为$\frac{1}{2}m\times\frac{1}{2}n$。

式算出所需存储容量：M＝974312位。选用EPROM芯片的集成度通常16K位、32K位……256K位等系列，一般尽量取集成度高的芯片，以减小体积和电源功耗，并可简化电路。设用256K位的芯片，则需用芯片数可按（2）式算出S＝4片（向上取整数）。如果有1兆位的芯片，则一片就够用了。

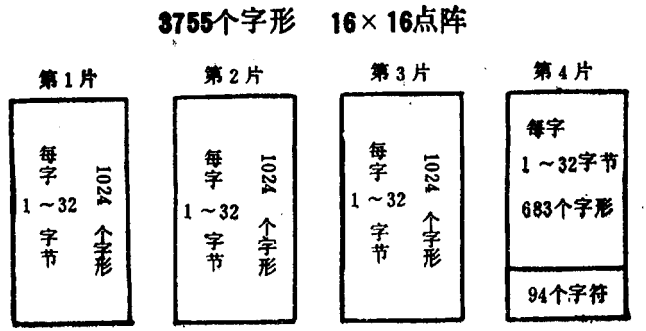
② 根据所选的芯片确定点阵数据的存放格式

如前所述，15×16点阵字形的数据量是每个汉字32字节，数据的排列次序是由上到下，第一行是第1、2字节，第二行是第3、4字节，以此类推，第16行是第31、32字节。以图4·9(a)的“啊”字为例，第1、2字节是00、00，第3、4字节是0F、7E，……第15、16字节是08、08。如果设计3755个汉字的固化字库，共有60080个字节，这些数据应怎样存放呢？

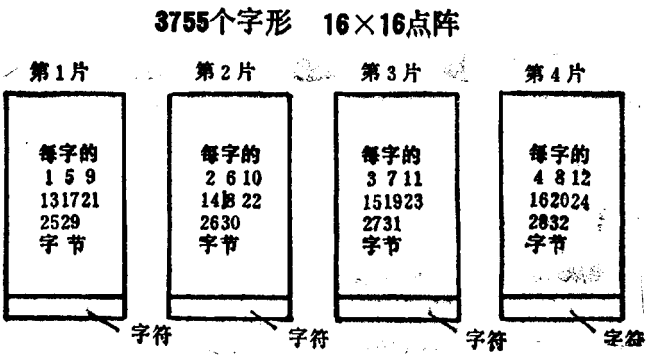
一种简单的存储方法是按照国标GB2312-80规定的编码顺序，将每个字的点阵数据依次集中存放在一块ROM芯片中，存满一块再存下一块。这种方式称为集中存放式。另一种存储方法是按照国标GB2312的编码顺序或其它特定的顺序，将每个字的点阵数据，以字节为单位分布存放在组成字库的若干块ROM芯片上，这种存放方式称为分布存放式。现以图4·12说明由4块256K位的芯片组成的字库两种不同的数据存放格式。

图（a）是集中存放格式，在第1～3片上各存放1024个15×16点阵的字形数据，第4片上存放683个字形数据，共3755汉字，余下为存放94个半角字符的字形点阵数据及空区。

图(b)是分布存放格式，第1片上存放3755个字的第1、5、9、……29字节；第2片上存放3755个字的第2、6、10、…30字节；第3片上存放第3、7、11、…31字节；第4片上存放第4、8、12、…32字节。每一片上余下的存储区存放94个字符及



(a) 集中存放式



(b) 分布存放式

图 4·12 固化字库的数据存放格式

作为空区。

集中存放格式的优点是字库排序简单，汉字检索软件有规律可循，硬件的设计也较为容易。但是，由于一个字形数据集中放置在一起，读取数据时按字节读取速度比较慢。分布存放

格式的优点是数据读取速度快，可靠性好，还可以提高存储空间的利用率，因此在用多块芯片组成字形发生器时，大多采用这种存放格式。当然，对于采用高集成度（兆位级）芯片的字形发生器来说，如果只用一块芯片就足以存放全部字形点阵数据，那当然是集中存放在一片上了。

③ 设计字形点阵数据并写入存储器

在完成以上两步以后，就可以设计字形点阵数据，按照格式将数据写入EPROM存储器。这是一项繁重复杂的工作。用人工设计点阵字形和读取点阵数据，可以按照上一小节介绍的方法进行，但具体做起来非常费时费事。因此，在标准点阵字模集和数据集制订公布以前，对点阵字形及其数据的设计，是通过应用专门的造字软件，以计算机辅助设计方法来实现。在国标GB5199、GB5007和GB6345公布实施以后，15×16、24×24、32×32三种常用的字形点阵数据都已有了统一的标准，除了需要增补的少量GB2312-80中未收的集外字以外，已基本上免除了设计点阵字形的麻烦，只需把标准点阵数据按照设计的排列格式写入EPROM即可。

（3）国标汉字点阵芯片的使用

在GB5007、5199和6345三项国家标准公布实施后，通过与国外技术合作，很快研制了用1兆位、2兆位、4兆位超大规模集成电路的MASK ROM国标汉字点阵芯片，并已大量生产供应市场。国标点阵芯片的使用，不但大大简化了字形发生器的设计过程，而且对降低汉字终端和汉字系统的生产成本、促进汉字数据处理和汉字字形识别等方面，都将有重要作用。为了便于使用，现将这三种标准芯片组的汉字点阵数据排列格式及有关内容介绍如下。

① GB5007芯片组

GB5007芯片组是24×24国标汉字点阵芯片组，共有三种型号：A型、B型和C型。

（A） GB5007A型

GB5007A型芯片组包括2块芯片，其中1块4Mbit芯片（型号5007A_1）和1块1Mbit芯片（型号5007A_2）组成国标基本集GB 2312汉字字形发生器。A型芯片组的汉字点阵数据排列格式如图4·13所示。

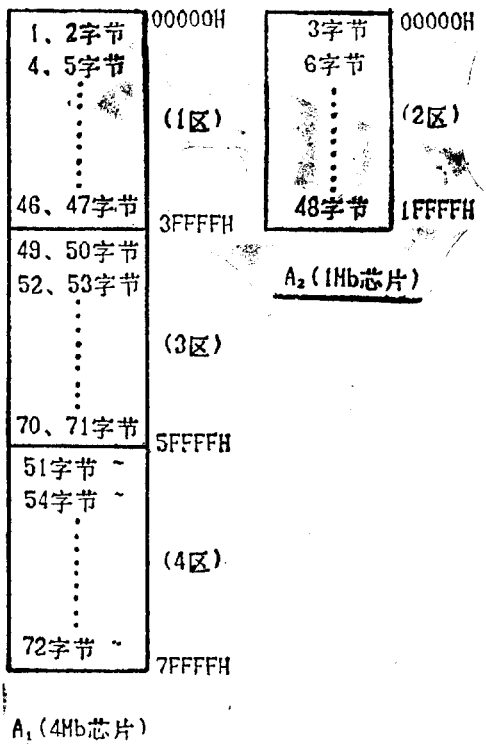


图 4·13 GB5007A型芯片组的数据排列格式

由图可见，A型芯片组的排列格式是经过专门设计的。它把A_1、A_2两块芯片的存储区统一划分成4区：A_1芯片的前2Mbit为1区，A_2芯片（1Mbit）为2区，A_1芯片的2～3Mbit为3区、3～4 Mbit为4区。汉字字形点阵数据以按区分布格式存放，即在同一区内，以国标码为序存放全部汉字字形的一部分点阵数据字节，但每一个汉字字形的72个字节则分布存放在4个区内。其分布规律是：72个字节中的1～48字节数据存放在1、2区，49～72字节数据存放在3、4区。各区内数据存放的分布规律是：字节数等于3的倍数的数据存放在2、4区，其中3、6、9…48字节存放在2区，51、54、57……72字节存放在4区；不是3的倍数字节的数据存放在1、3区，其中1区是2Mbit，存放1、2、4、5、……46、47共32字节，3区是1Mbit，存放49、50、52、53、……70、71共16字节。

按照GB5007A型芯片组的数据排列格式，可使字形发生器的外围电路和数据检索程序的设计得到简化。只需在专用的解码门阵列集成电路芯片B202的支持下，组成24×24国标汉字点阵字形发生器。它存放了包括国标基本集GB2312的全部汉字和字符图形外，还充分利用存储空间，增补了集外汉字122个及大量字符，增补的主要内容可参阅表4-2、表4-3和表4-4。

(B) GB5007B型

GB5007B型芯片组包括3块芯片，其中2块2Mbit芯片（型号5007B_1、B_2），1块1Mbit芯片（型号5007B_3）组成国标基本集GB2312汉字字形发生器。B型芯片组的汉字点阵数据排列格式如图4·14所示。

由图可知，三块芯片分成2区，其中两块2Mbit芯片（B_1

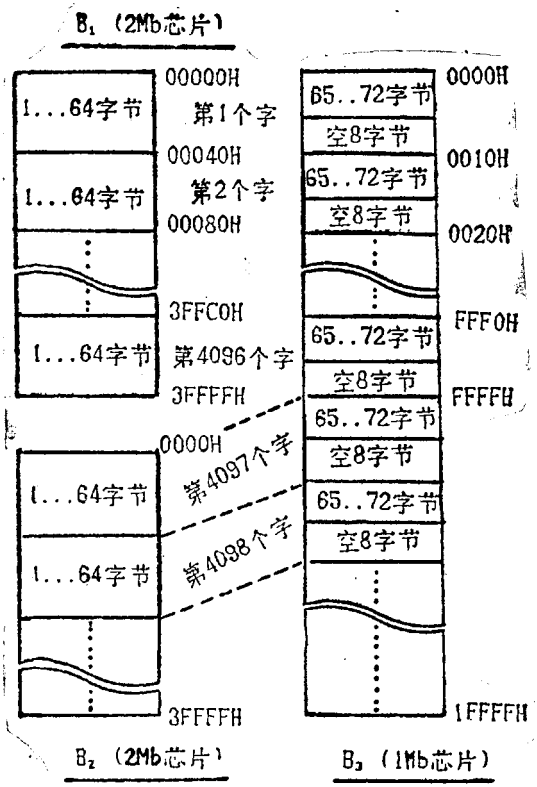


图 4·14 GB5007B型芯片组的数据排列格式

和B_2）为1区，存放字形点阵数据第1至64字节，1Mbit芯片B_3为2区，存放字形点阵数据第65至72字节，为了与2区对应以便检索，每存放8字节后空8字节。B_1芯片存满4096个汉字的第1至64字节数据后，从第4097个汉字开始存放在B_2芯片上，在B_3芯片则从9000H开始存放第4097个汉字的后8字节数据。必须注意，以上注明的芯片存储单元的16进制地址数，是

以“字节”计的。

GB5007 B型芯片字库中汉字和字符图形的排列顺序是：一级字、二级字在前，字符图形在后，共存放了国标基本集中的84个区，即一、二级汉字从16至87区，共72个区；字符图形从1区至12区，共12个区，其中10区至12区是增补区。增补的主要内容有：

• 在08区78—94位和11区70—94位共增补汉字42个。

• 在10区增补94个第3区半角字符。

• 在12区增补94个第3区18×24字符，可保证宽行打印机一行输出132～136个字符。

• 在11区35至69位增补35个半角控制符。

• 在11区01至32位增补32个半角拼音符。

增补汉字、字符的具体内容及区位可参阅该芯片的技术资料。

(C) GB5007C型

GB5007C型只用一块4Mbit芯片（型号GB5007C），放下了包括GB2312-80的全部一级、二级汉字，字符部分缺少第4至8区，保留第1、2、3、9区，增加第10区为半角字符区，其排列顺序是：一级字、二级字共72个区，字符4个区，再加半角字符区，共存放77个区。此外，还在各区空位上增补以下内容：

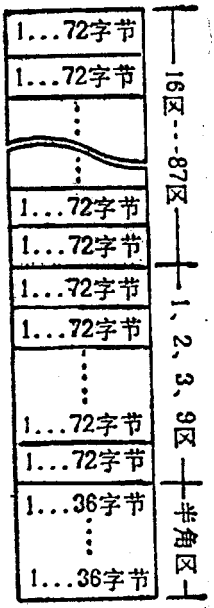
• 在2区01～16位增补16个纵书符。

• 在9区01位增补一个半角符$。

• 在9区80～94位增补15个制表符。

GB5007C型芯片的存放格式完全按照国标GB2312的编码顺序，在一块芯片上集中存放，如图4·15所示。

② GB5199芯片



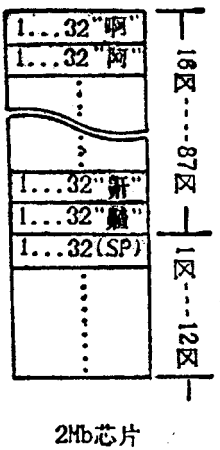


图 4·15 GB5007C型芯片数据存放格式　　图 4·16 GB5199芯片数据排列格式

GB5199字库芯片有A、B两种型号，都只用一块2Mbit芯片，可以提供15×16点阵字形输出。

GB5199两种型号芯片的数据排列格式都是将同一字形的32个字节数据集中顺序存放的，如图4·16所示。

但是，字的排列次序不同。A型芯片的排列次序经过精心安排，需要专用解码电码芯片B202的支持；B型芯片的排列次序以国标GB2312为序，一、二级汉字在前，字符在后。除了国

标GB2312中已收的汉字字符外，这两种型号的芯片也增补了一些汉字和字符，增补的主要内容，A型芯片与GB5007 A型芯片大体上相同，B型芯片与GB5007B型芯片大体上相同，故不再一一列出。

③ GB6345芯片组

GB6345芯片组是32×32国标汉字点阵芯片组。它是两块4Mbit芯片，在专用的解码门阵列集成电路芯片B202支持下，组成高质量汉字字形发生器，供针式打印机及激光、液晶打印机输出汉字32×32点阵字形。

芯片存放数据的排列格式如图4·17所示。

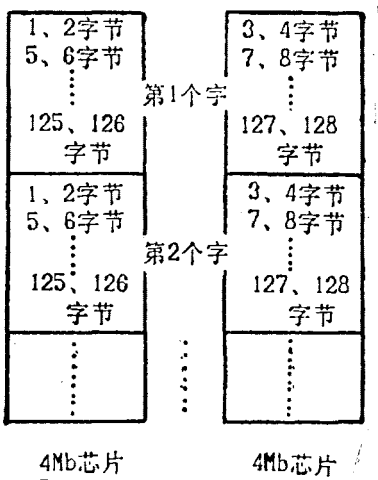


图 4·17 GB6345芯片数据排列格式

由图可见，GB6345芯片以分布方式存放国标GB2312中的汉字和字符，在2块芯片中分别存放同一字形的64个字节字形数据。数据是以“字”为单位，1字等于2字节，奇数“字”

在左，偶数“字”在右。该芯片还增补了GB2312集外的汉字和字符图形，增补的具体内容及区位安排与GB5007A型芯片完全相同。

（4）字形发生器与主机的连接

在汉字终端和汉字微型机系统中，用半导体存储器组成的字形发生器（即固化字库）与主机（即主CPU）的连接，大致有三种方式：内存连接方式、I/O连接方式和插接兼容（即汉卡）方式。

① 内存连接方式

内存连接方式是把字形发生器的ROM芯片，如同其它内存储器一样，与主机CPU的系统总线直接连接，如图4·18所示。

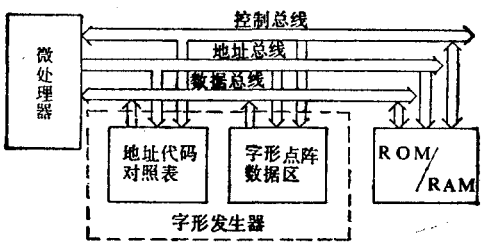


图 4·18 字形发生器的内存连接方式

图中的字形发生器由两组ROM组成，一组存放字形点阵数据，另一组存放汉字代码与点阵数据地址对照表，都按照内存统一编址存放。假设某汉字终端的字形发生器容量32K字节，需占用内存地址为0000～7FFFH，则系统的其它内存地址单元应从8000H开始编址，不能重复。

内存连接方式的字形发生器操作过程比较简单，当需读取

某汉字字形时，先取出该字代码，在代码/地址对照表中查出对应的内存地址；然后将该地址送入字形点阵数据存储区，就可读出该字的字形点阵数据。但是，内存连接方式的字形发生器占用主机的内存地址空间过多，只能用于字数不多的小型设备上，通用的字形发生器一般不采用这种连接方式。

② I/O连接方式

I/O 连接方式的字形发生器是把ROM芯片和外围控制电路一起，通过并行的I/O口（即输入/输出口）连接到主机的系统总线。这种连接方式，使字形发生器只占用系统的一个外部设备地址和代码/地址对照表的少量内存空间，能够大大节省系统内存。由于使用半导体存储器芯片，数据的读取速度与内存方式相差无几，因此应用比较普遍。图4·19是一种字形发生器I/O 连接方式的示意图。

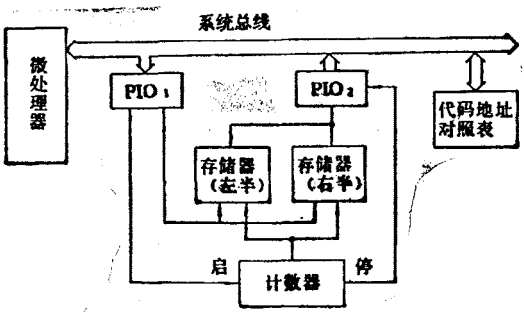


图 4·19 字形发生器I/O 连接方式

图中作为字形发生器主体的两组存储器，通过并行I/O 接口芯片PIO与系统总线相连。PIO有两个端口 PIO_1 和 PIO_2，PIO_1口用来接收数据，PIO_2口用来发送数据。两组存储器分别存放字形点阵的奇数字节和偶数字节。

I/O 连接方式的字形发生器操作过程：主 CPU 读取在内存某地址的一个汉字代码，从内存的代码/地址对照表中查到该字在字形发生器存储单元的首地址，把它送入 PIO_1 口，并启动计数器。计数器的作用是控制读出点阵数据的字节数。例如对24×24点阵字形，计数器置72（48H）。CPU通过 PIO_2 口在左半右半存储器中读取点阵数据，当读满72个字节后，计数器送出停止信号，PIO_2口停止输出。I/O连接方式的字形发生器种类较多，具体操作过程也不尽相同，这里仅举例说明它的基本原理。

③ 插接兼容方式

插接兼容方式连接的字形发生器又称为“汉卡”，这是目前广为流行的字形发生器与主机的连接方式。一台通用的微型机或输出设备，如显示器、打印机等，通过插入“汉卡”，不需变动任何硬件就可以具有汉字处理或汉字输入输出功能，成为一台汉字微型机或汉字显示器、汉字打印机，使用十分方便，因而得到了广泛的应用。

所谓“汉卡”就是装有半导体字形发生器及控制电路的一块标准总线插件板。当把汉卡插入系统总线槽口以后，系统向它输入任一汉字代码，就可以用直接存储器存取(简称DMA)技术，方便地读取汉卡中存储的字形点阵数据。

汉卡的种类很多，以各种型号微机总线结构不同来分，有Apple汉卡、IBM汉卡等；以适用的不同设备来分，有打印机汉卡、显示器汉卡等。现以一种带CPU的微机汉卡为例说明其基本工作原理，这种汉卡的结构框图如图4·20所示。

由图可见，汉卡的主要部分是在 CPU 通过管理程序控制下的字形发生器。监控程序通过并行接口与主机 CPU相连，它接收来自主机的汉字代码，在汉卡CPU控制下把汉字代码送

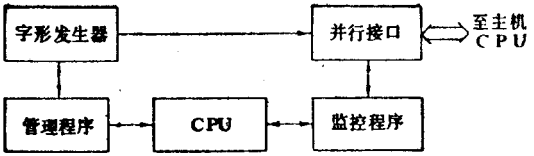


图 4·20 汉卡的结构框图

入管理程序，在字形发生器中找到相应的汉字字形点阵数据，直接送并行接口输出。

汉卡在监控程序和管理程序的分析处理下，可以接收来自主机的各种格式的汉字代码，将它们转换成国标码，对字形发生器进行字形点阵数据检索。因此，对于各种不同的汉字输入编码，只要在字库管理程序中设有编码方案的相应软件，主机CPU就可以向汉卡直接送入该方案的汉字输入代码，并得到点阵数据输出。这种汉卡往往就以该输入编码方案或输入法命名，例如仓颉汉卡、联想汉卡、五笔字型汉卡等等。设计优良的汉卡，不但能适应多种输入编码，而且还有其它多种功能。

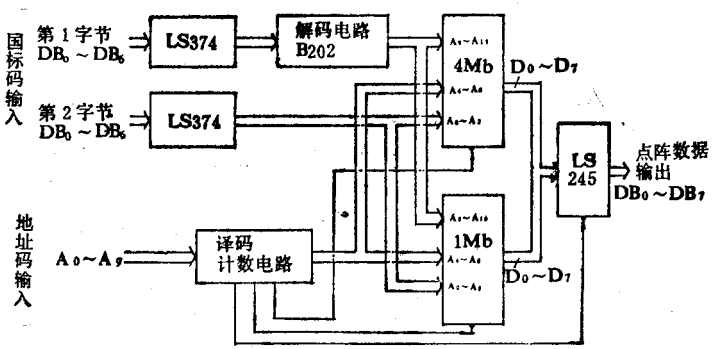


图 4·21 4Mbit 24×24点阵汉卡原理图

如造字及汉字编辑功能、多种字形数据变倍输出功能等。由于采用1Mbit～4Mbit的MASK ROM芯片，使汉卡的硬件设计大为简化。图4·21所示是一种实用的4Mbit芯片 24×24 点阵汉卡原理图。

4·3 汉字显示器

1. 概述

汉字显示器是由显示器件和有关的控制与驱动电路组成的设备。它能把存储在机器内部的汉字字形信息（电信号）转变成能被人的视觉感受到的汉字字形（光信号），并以一定的格式显示输出。

在汉字终端和汉字微型机系统中，汉字显示器除了作为汉字和字符图形的输出显示外，还用于汉字输入编码时的输入显示、重码字的二次选择显示、文字编辑增、删、改显示等人机对话方面，是实现汉字信息输入、输出的主要设备之一。与印字输出方式相比，显示输出方式的响应速度快、无机械磨损和噪声，设备可靠性好，使用方便。因此应用十分广泛。

（1）显示器的分类和特点

显示器是一种将电信号转变成光信号的设备，它所利用的基本物理现象是电光现象。常见的电光现象很多，按照它们不同的发光原理，可分为两大类，即电致发光和电光效应。

电致发光是将电能直接变换成光能的电光现象，例如：气体放电发光、阴极射线发光、电场发光等等，都是常见的电致发光。应用电致发光原理做成的显示器件是一种有源显示器件，或称主动型显示器件。有源显示器件的种类很多，最常见

的电子束管（即阴极射线管）显示器件，已普遍应用于电视机、示波器等显示设备中。此外，还有场致发光显示器件，等离子体显示器件以及发光二极管显示器件等，也日益得到广泛的应用。

有源显示器件的共同特点是显示的清晰度好。因为光能是显示器件本身发出的，与使用环境的光亮程度关系不大。

电光效应是通过外加电能后使某种物质的光学性质发生变化的电光现象。常见的电光效应如晶体的普克尔效应、液晶的动态散射效应、铁电陶瓷的电控双折射效应等，它们都是在物质被通电后，改变了通电部分对光的折射率、反射率或透射率，便可显见出文字或图形。应用电光效应原理做成的显示器件是一种无源显示器件，或称被动型显示器件。常见的无源显示器件有液晶显示器件、各类光阀等。由于它们本身并不发光，需要在外光源照射的条件下，靠不同的折射（或反射）率才能显现图文，因此与使用环境的光亮程度有很大关系。

按显示器件的不同结构，汉字显示器可分为电子束管显示器和平板型显示器两大类。

电子束管显示器是发展最早，技术最成熟的显示设备。它是在一个真空容器里，用电子束轰击荧光粉发光，再用适当的控制和驱动电路控制发光点在荧光屏上的位置、亮度和颜色，便能形成文字、图形和图象的显示。电子束管显示器的分辨率高，响应速度快，便于显示彩色图文。随着微型计算机技术进入显示领域，使电子束管显示器既具有存储、显示的功能，又能在输入设备（如键盘、光笔等）的配合下对显示内容实行控制、干预和人机对话，是目前应用最广的显示设备，在当前和今后较长时期内，电子束管显示器仍将在显示领域中占有主导地位。电子束管显示器的缺点，一是体积大，由于“电子枪”

的结构所限，使它总是拖着一根长尾巴；二是工作电压很高（几千伏）；三是易碎、不耐震。随着终端设备向小型化、低电压、低功耗方向发展的潮流，部分电子束管显示器将被平板型显示器所取代。

平板型显示器是近一、二十年迅速发展起来并且有着广阔应用前景的新型显示设备。它是用两块玻璃平板夹持住具有电光特性的物质构成的。这些物质有荧光粉、惰性气体、液晶等多种。在两块玻璃平板上分别做上了许多条平行的横向电极和纵向电极（即X、Y电极）。在X、Y电极空间正交处的发光物质形成一个个象素点矩阵。可以通过X、Y电极来选择不同的象素，并用信号电压控制某些象素发光（或改变光特性），就能实现图文显示。

平板型显示器的结构简单，体积小，工作电压低，最高不过几十伏至一、二百伏，有的还可直接用集成电路驱动，有利于终端设备的小型化。平板型显示器的缺点，一是显示的响应速度慢；二是分辨率低，显示多级灰度比较困难。目前主要用在计算机及终端设备的一般图形文字显示方面，对高分辨率的活动图象显示，特别是彩色图象显示还比较困难。

现将两类显示器件的性能特点及应用范围列表见表4·9。

（2）汉字终端对显示器的性能要求

汉字显示器是一种计算机字符、图形显示器。它既要显示原有的字母、数字及各种符号，还要能显示比较复杂的汉字字形及图形符号。因此，汉字显示器除了应满足一般字符显示器的亮度、对比度等基本要求外，还有一些特殊的要求，如分辨率、每屏显示字数、显示速度以及余辉时间等，现分别讨论如下。

表4·9　　两类显示器件比较表

	电子束管显示器件	平板型显示器件
典型器件	显象管、示波管、存储管、投影管	场致发光板、等离子体显示板、液晶显示板
结构原理	用电子枪扫描荧光屏，通过控制驱动电路显示信息	用两块玻璃平板夹持工作物质，通过控制驱动电路显示
主要特性	是一种电真空器件，显示响应速度快，分辨率高，技术成熟可实现多级灰度及彩色显示	是一种固体平板器件，体积小，重量轻，工作电压低，驱动容易，可做成大面积显示
缺　点	工作电压高，体积大，笨重，易碎，显示面积不能太大	分辨率较低，响应速度较慢，交叉效应未能彻底消除
应用范围	广泛应用于各种显示领域	主要用于显示文字和图形

① 分辨率

分辨率又称为分辨度，是显示器件对显示图形和文字细微部分的分辨能力，这是标志显示器质量高低的一项重要指标。

分辨率的基本定量概念是指能分辨的单位长度内最多线条数（如每毫米若干线）或整个幅面的相对分辨能力。对于电子束管显示器件来说，常用在一定显示面积上的扫描线数来表示分辨率。例如电视显象管在垂直方向的扫描线数为625行，这就是它的垂直分辨率。并由此可以推算它的水平分辨率，因为图象显示的标准宽高比为$\frac{4}{3}$，水平分辨率就是$625\times\frac{4}{3}=834$点。整个幅面的分辨能力为$625\times834=521250$点。

在字符图形显示中，分辨率高低也可用光点的粗细来衡量。例如字符显象管的一个水平行能显示300个亮点和300个暗

点，它的分辨率就是600线。显然，光点直径愈细，显象管的分辨率愈高。这对于图文显示是有利的，因为分辨率的提高能增加屏幕显示的每帧字数，改善字形质量。但对电视图象显示来讲，光点过细收看的效果不一定更好。

平板型显示器的分辨率通常用单位长度内的电极线数表示。电极线数愈多，分辨率愈高。但是由于交叉效应的影响，平板显示器的分辨率不易做得很高。

② 每屏显示字数及显示速度

汉字显示器的每屏显示字数及显示内容的切换速度，是衡量显示器性能的又一重要指标。

在显示器屏幕尺寸一定的条件下，每屏显示字数的多少，取决于显示器的分辨率以及所需显示字形的精密度。例如，用分辨率为640×640点阵的显示器来显示24×24点阵字形。以字与字间的距离为一个“点”，行与行间的距离为10个“点”计算，则每屏最大显示汉字数约为25×16＝400个。

每屏显示的汉字及字符数较多，有利于输入文稿的编辑与应用。但屏显字数的增加会影响屏显内容的切换速度。因为屏显切换速度取决于电子束的扫描周期。要保证在扫描周期的正程内有更多的字形信息输出，在扫描频率一定的条件下，就得延长扫描周期。否则，有些字形信息将会落入扫描周期逆程而无法得到正常显示输出。

③ 余辉时间

余辉是荧光粉等发光物质在电子束轰击停止后仍然持续发光的一种物理现象。通常以电子束停止轰击到荧光粉发光强度衰减到初始值1％的时间，称为余辉时间。

余辉时间的长短，在不同的使用场合有不同的要求。例如，电视显象管的余辉时间不能太长，否则在观看球类比赛时

会看到球的后面拖着一个“尾巴”。作为图文显示用的显象管的余辉时间可以长一些。这样可以适当延长显示刷新周期。这对于降低显示器的视频信号频率、减少显示屏幕的闪烁现象是有好处的。

不同发光材料的余辉时间是长短不一的。通常把余辉时间分成三类：小于1毫秒的称为短余辉时间；1～100毫秒的称为中余辉时间；大于100毫秒的称为长余辉时间。汉字终端一般采用中、长余辉时间的显示器。

④ 光点颜色

显示器显示光点的颜色，随着发光物质的不同材料和制造工艺而异。不同的发光材料在电子束轰击下可以发出不同波长的可见光，因而能得到不同颜色的显示光栅。在电子束管显示器件中，通常采用硫化物荧光粉作为发光材料，并在其中掺入少量称为激活剂的杂质。通过改变激活剂的数量，就能比较方便地改变荧光粉的发光颜色。在黑白电视显象管中要求发白色光；在字符显象管中通常要求发绿色光。一般认为绿光对人眼的刺激性较小，能减少操作者长时间工作时的眼睛疲劳。

具有彩色显示性能的汉字显示器，能使显示文字信息的内容分门别类、突出重点，因而更便于使用，但彩色显示器的结构复杂，价格较贵。

除了以上各项要求外，高质量的汉字显示器还能实现多功能的窗口显示和多字体多字号的显示，使屏幕上同时显示几个独立的信息内容，几种不同字体和不同尺寸的字形，这对于文件编辑处理，提高机器操作效率十分有利。此外，对显示设备的小型化、低功耗、耐震动和耐高低温等要求，也是不可忽视的。

2. 阴极射线显示管的基本工作原理

阴极射线管显示器是一种应用最广的电子束管显示器，可以用来显示各种复杂的文字图形以及活动的图象。它的基本工作原理是：先把阴极加热使它能发射电子，再用电磁力把阴极发射的电子聚焦成束，然后在外加同步信号和图文信号（合称为视频信号）的作用下，通过偏转系统来控制电子束的运动方向，去轰击荧光屏的不同象素使之发光，就能把外加图文信号所代表的图形文字显示出来。

（1）阴极射线管的结构

阴极射线管显示器又称为CRT显示器，它的基本部件是一个阴极射线管。图4·22是阴极射线管结构示意图。

由图可见，阴极射线管主要由电子枪、偏转系统和荧光屏三部分组成。

电子枪是电子束的发射体。它的作用是在通电加热之后产生强度足够大、而且能调节的电子流；它还能将电子流聚焦成束，使打到荧光屏上的电子束直径足够小。

偏转系统是阴极射线管的主要控制部件。没有偏转系统，电子枪发出的电子束只能在荧光屏上打出一个小亮点。只有通过偏转系统的控制作用，才能使电子束犹如一支灵巧的画笔，在屏幕上描绘出种种变化无穷的图象和文字。

根据不同的控制作用原理来分，偏转系统有静电偏转和磁偏转两种方式。静电偏转系统是在管内电子束行进的路线上装一对静电偏转板，通过改变板上电压大小来改变电子束的运动方向，达到控制的目的。静电偏转系统的结构简单，能适应外加高频控制信号。但偏转的角度较小，一般只能做到在25°以内，

故要增大显示屏幕的面积比较困难。

磁偏转系统是用装在管颈外面的偏转线圈产生的偏转磁场来控制管内电子束运动方向。图4·22所示的阴极射线管就是采用磁偏转系统，它有两对偏转线圈，垂直的一对和水平的一对互相重叠在一起，套在管颈上。两对偏转线圈用来分别控制电子束的垂直和水平方向运动的偏转角度，最大偏转角可达114°。通常用在要求显示面积较大、图象质量较高的场合。由于线圈的电感和分布电容量较大，磁偏转系统不适宜用于需外加高频控制信号的场合。

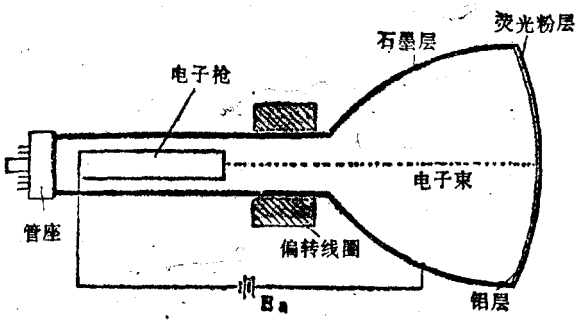


图 4·22　阴极射线管结构示意图

荧光屏好比一张能重复使用的“画图纸”，可供“电子画笔”（电子束）在上面不断画出复杂、美观的图象和文字。荧光屏的主要成份是在一种特制的烟灰玻璃屏上涂以荧光粉，再在荧光粉上蒸一层极薄的铝膜，并把它与管壁的石墨导电层连在一起。

荧光粉是一种受到电子束轰击后能发光的物质，通常是硫化锌类的化合物，再掺加少量银、铜一类的激活剂，以增加发光亮度，并可以用来改变发光的颜色。荧光粉在受到高速电子轰

击下，还会产生二次电子发射，即本身又发射出电子。二次发射电子的不断积累，就会排斥后来电子束的轰击，影响屏幕的正常显示。因此必须有一层导电的薄铝膜，把二次电子排除掉。由于铝膜的厚度只有0.05～0.5微米左右，对荧光屏的正常发光没有影响。铝膜还可以防止质量较大的负离子对荧光屏的轰击，以免荧光屏的中心产生黑色的离子斑。

（2）光栅扫描及字形显示

如前所述，CRT荧光屏上显示的图象，是在偏转系统作用下由电子束扫描而成的。其扫描过程如图4·23所示。

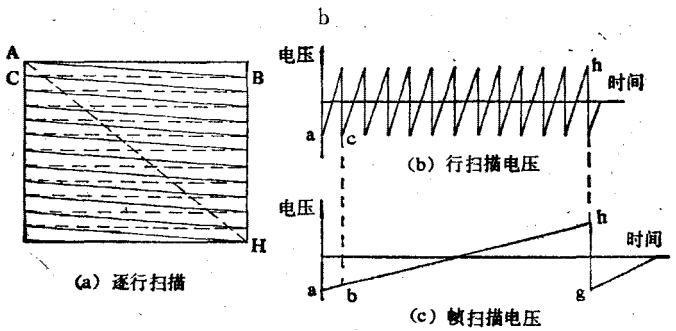


图 4·23 电子束扫描过程示意图

图(a)是电子束在屏幕上的逐行扫描光栅图。它是由加在水平偏转线圈上的锯齿波电压，即如图（b）所示的行扫描电压，与加在垂直偏转线圈上的锯齿波电压，即如图（c）所示的帧扫描电压共同作用的结果。

见图（b），在行扫描电压abc一个周期作用下，电子束从左到右地扫描，并在荧光屏上产生一条水平亮线，即图（a）中的AB线段。然后在电子枪的调制极上加负偏压（称为消隐电压），故在BC段荧光屏不发光。AB段称为扫描的正向行

程，BC段称为逆向行程。在帧扫描电压bh段的作用下，见图（c），使水平扫描线段逐行下斜，直至右下角H点，就完成了对整个屏幕的帧扫描。当扫描完1帧以后，在帧扫描电压hg段的作用下，电子束回到起始点A，并接着开始第2帧的扫描过程。

实验证明，当光的“亮”与“暗”交替变化在每秒30次以上时，人的视觉就不会感到光的闪烁现象，这是由于人眼的“视觉暂留”作用的缘故。以上所述的荧光屏的发光现象，只是当电子束轰击到某个象素点时，该点才发光。电子束扫描过后就不发光了。但是，当扫描的速度足够快时，虽然该象素是在亮、暗交替变化，但看起来好象是一直亮着。因此，只要当把图4·23（c）的帧扫描电压的频率（即帧频）提高到30Hz以上，就会在荧光屏上呈现出稳定的光栅，不会使人感到闪烁了。

行扫描电压的频率（即行频）高低，取决于一幅画面的扫描行数及帧频数。我国电视制式规定扫描行数为625行，采用隔行扫描方式，即1帧分成2场扫描：一场扫描奇数行，一场扫描偶数行，因此可使帧频频率降低$\frac{1}{2}$。在字符显示器中，则多采用图4·23（a）所示的逐行扫描方式。

例如，某一字符显示器每帧画面为625行，每秒扫描50帧，即帧频＝50Hz，逐行扫描时的行频为$625\times50=31250$Hz。即每行扫描时间为$\frac{1}{31250}$秒＝32微秒。其中正程时间为26微秒，逆程时间为6微秒。

汉字显示器是一种字符图形显示器，它所显示的汉字字形和其它字符图形，通常是用点阵方式表示的。显示器的扫描制

式不受电视显象管扫描制式的约束，可以根据具体条件设计。例如每帧画面水平扫描行数的确定，可根据满屏显示的汉字数及字形的点阵数算出，不一定是625行。

假设某汉字终端的显示器设计每帧最多可显示15×16点阵的汉字640个。采用逐行扫描方式。计算该显示器的水平扫描行数、分辨率以及扫描电压频率。

首先确定显示器的屏幕显示格式。将640个汉字分成16行、每行40个字排列，就可安排满屏汉字的显示格式，如图4·24所示。

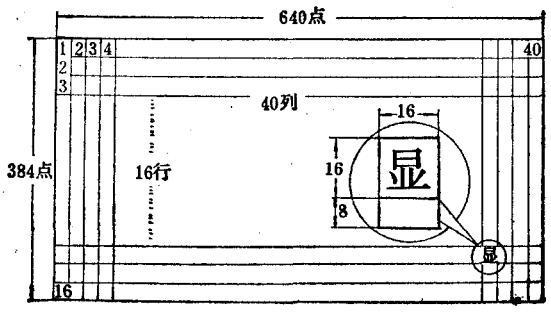


图 4·24 汉字字形的屏显格式

由图可见，在屏幕的有效显示面积内，划分16行×24列的小方块矩阵，若规定字与字间留出1个“点”间距，行与行之间留出8个“点”间距，则一个小方块的点阵数应为16×24，即宽16点，高24点，于是可算出水平扫描行数应为：

$$24 \times 16 = 384\text{行}$$

还可算出显示器的水平分辨率，即在水平扫描线一行内的分辨率应为：　　$16 \times 40 = 640$点

按要求采用逐行扫描方式，设帧频为30Hz，可算出行扫

描频率应为：　　$384 \times 30 = 11520\text{Hz}$

将与上述扫描频率同步的数字脉冲信号加在电子枪的调制极上，当信号为“0”时，屏幕上是一个暗点；当信号为“1”时，屏幕上出现一个亮点。于是在光栅扫描过程中可使这些暗点和亮点稳定地重现。显然，如果把来自字形发生器的点阵数据与扫描频率同步后加到电子枪的调制极，就会在屏幕上显示出相应的点阵字形。而且，可以通过对加于显示器输入端字形点阵数据排列顺序的控制，确定每个字形在屏幕上所处的位置，按照要求的汉字文本格式输出。

应当指出，CRT显示器的显示扫描方式不是一个字形接着一个字形扫描，而是逐点行扫描方式，即对同一字行中的字形，每次扫描出它们的同一点行。对于15×16点阵字形，必须连续扫描16个点行，才能把同一字行中的字形全部显示出来。因此，应当按照这个特点适当安排加在显示器输入端的串行数据序列，才能达到正确显示的目的。

如前所述，字形发生器存放的汉字点阵数据多为横排格式，如图4·25所示，是24×24点阵字形的横排数据格式。

图中的1…72是字形点阵数据的字节数。$D_0 \cdots D_7$是每个字节的信息位数，低位在左，高位在右，与从左至右的扫描方向一致。因此，当显示一行汉字时，在每一扫描行中，包括本行每个汉字字形点阵的3字节数据，并按从左到右，由低位到高位的次序排列。于是可得到一行40个汉字的2880字节数据排序为：

第1字的第1、2、3字节，第2字的第1、2、3字节……第40字的第1、2、3字节；第1字的第4、5、6字节，第2字的第4、5、6字节……第40字的第4、5、6字节；……直至本字行的最后一个扫描行：第1字的第70、71、

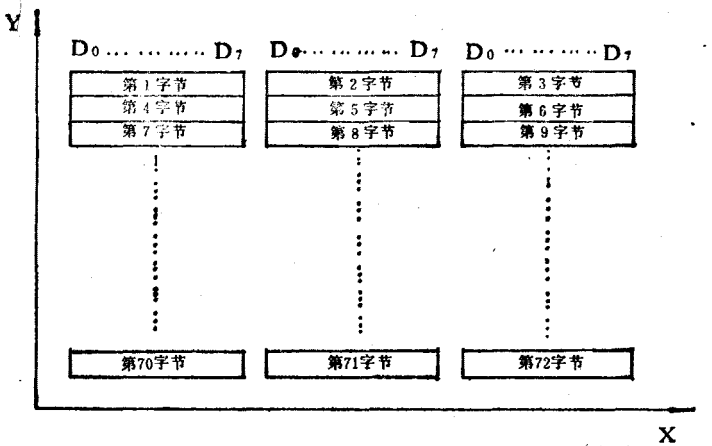


图 4·25　24×24字形点阵数据的横排格式

72字节，第2字的第70、71、72字节……第40字的第70、71、72字节。以上合计有24个扫描行，每个扫描行有3×40=120字节。

3. 平板型显示器的基本工作原理

平板型显示器是近十多年迅速发展起来并在小型和移动式汉字终端设备中得到广泛应用的新型显示器件。它的基本工作原理，是在分别刻有X、Y电极的两块平板玻璃上，加上构成字形点阵的适当电压，促使X、Y电极不同交叉点的象素发光或者改变光学特性，来实现字形信息的显示输出。

根据两块玻璃平板所夹持的不同物质来分，平板型显示器有液晶显示器、等离子体显示器和场致发光显示器等数种。

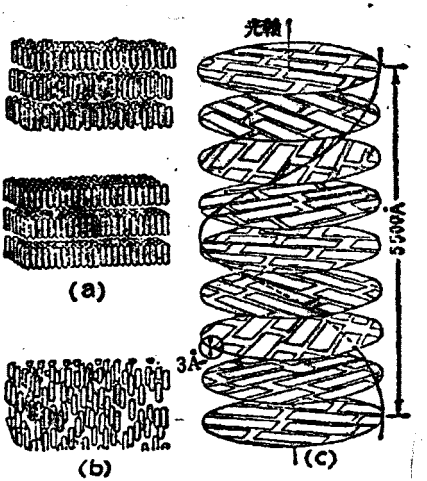
（1）液晶显示器

液晶显示器简称LCD，是应用最广泛的一种平板型显示

器。

液晶是“液态晶体”的简称。顾名思义，液晶既有液体的流动性，又有晶体各向异性的特点，具有很规则的组织结构。目前，人们已经发现了两千种液晶材料，根据分子排列的不同结构，大致可分为三种类型：近晶型、向列型和胆甾(读“灾”)型三种。近晶型液晶的分子排列整齐，恰似盛在盒子里的火柴；向列型液晶中的细长分子大致也按同一方向排列，但不如近晶型那样整齐，层次结构不明显；胆甾型液晶是多层结构，每层中分子排列与向列型类似，但各层分子排列方向会产生旋转，各层的取向呈周期性变化的螺旋状结构。液晶显示器主要使用后两种液晶。

液晶的三种不同分子结构如图4·26所示。



(a) 近晶型；(b) 向列型；(c) 胆甾型

图 4·26 液晶的分子结构

① 液晶的显示原理

如上所述，液晶是一种分子结构排列整齐的有机透明体。液晶由于液体的流动性，没有固定的形状，液体中的分子集团可以自由移动。液晶的各向异性主要表现为：在平行分子轴方向和垂直分子轴方向上，它们的导电率、介电常数和折射率均不相同。液晶的特殊结构使它具有这样的特殊性质：用电、热、磁和机械力等外来因素均可改变液晶的物理状态。在液晶显示器中，主要是应用在电压作用下的液晶动态散射效应来实现显示的。

所谓动态散射效应，就是在外加电压作用下，使透明的液晶变得浑浊不清的一种现象。图4·27所示是一种向列型液晶的动态散射效应示意图，这类液晶的分子是长条形的，分子在集团中可以排列成两种状态：一种是平行取向状态，即分子排列

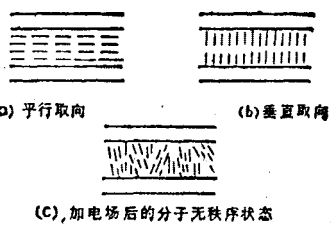


图 4·27 液晶的动态散射效应

方向与显示板平面平行，如图（a）所示。另一种是垂直取向状态，即分子排列方向与显示板平面垂直，如图（b）所示。当在某一个象素单元上施加低频电压时，使液晶内部产生许多带电粒子，由于带电粒子的碰撞作用，打乱了液晶分子原来的整齐排列次序，如图（c）所示，使液晶变得浑浊，在外来光的照射下，由于不同的折射率使光发生散射，因而显示出字形或图象来。

当去掉外加电压以后，液晶分子又会逐渐恢复原来的排列状态，液晶就变得透明，字形自行消失。

目前，液晶的彩色显示性能引起人们的广泛注意，它主要是应用液晶的电场变色效应原理实现彩色显示。如胆甾型液晶分子层取向为螺旋状，在外加电场作用下，螺距产生变化，使它对光的吸收作用随之变化，从而液晶的颜色也发生变化。例如，有一种胆甾型液晶，不加电场时呈红色，加6伏/微米电场时呈黄色，加12伏/微米电场时呈绿色，加16伏/微米电场时则呈蓝色。当选通电路给交叉点的象素单元加上不同的电场后，液晶显示板便可呈现由各种不同颜色组成的字形或图象。

② 液晶显示板的驱动电路

液晶显示板的一个显著特点是耗电量很小，以有效显示面积计算，一般约为0.1mW/cm²；它的工作电压比较低，一般为几伏到三、四十伏，故可以直接用集成电路驱动，使用非常方便。图4·28所示是一种具有动态散射效应的向列型液晶显示板驱动电路原理图。

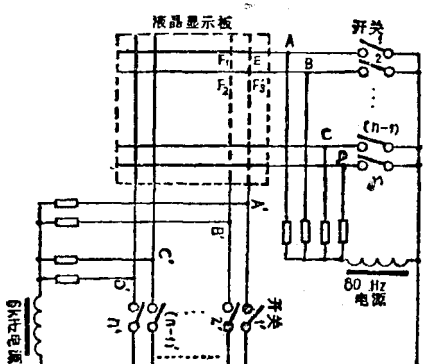


图 4·28 液晶显示板的驱动电路原理图

图中左上角是一块液晶显示板，它具有n条横向电极（X电极）和n′条纵向电极（Y电极）交叉构成的n×n′个象素单元。每一电极通过电阻与电源相连，并由开关1…n和1′…n′控制接地。液晶显示板的每一象素均用80Hz的低频电压进行书写（即显示），用6kHz的高频电压进行擦除。通过开关1…n和1′…n′的通断控制，便可对显示板每个象素进行控制，并显示信息。

例如，要显示象素E时，将开关1断开，开关1′接通。这时，在80Hz书写电压的作用下，象素E显示。与此同时，开关2…(n－1)、n均接通，开关2′…(n－1)′、n′均断开，显示板上除E以外的其它象素都在6kHz的擦除电压作用下被清除显示，图中的开关实际是由逻辑集成电路控制的。

液晶显示器的驱动电路比较简单，功率消耗很少，对减小设备体积非常有利。但是，液晶是一种被动型显示器件，本身不会发光，只有在光照的条件下才能看到显示图文，故在黑暗的环境中不能使用。

（2）等离子体显示器

等离子体显示器简称PDD。它是以等离子体（或泛指为气体放电）发光或激发荧光粉发光原理制成的平板型显示器件。

按工作原理来分，等离子体显示器有交流型和直流型两大类。汉字显示主要用交流型等离子体显示器。

① 交流型等离子体显示板基本结构

交流型等离子体显示板的基本结构如图4·29所示。在上、下两块玻璃基板上各敷有多条平行细电极（X、Y电极）及介质层，介质层上有一层能抗离子溅射、而且次级电子发射系数高的保护层，然后以低熔点玻璃粉封接，形成空间相距约150

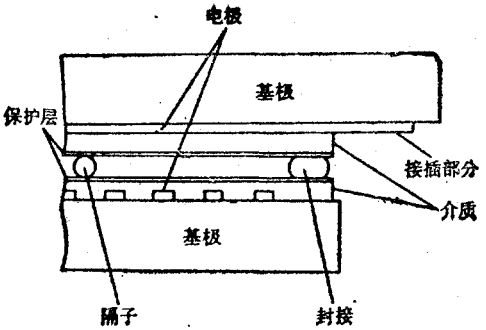


图 4·29　交流型等离子体显示板结构示意图

微米的真空密封腔体，腔体内充入氖、氩等惰性气体。基板上的X、Y电极在空间相互正交，每一对正交电极的空间都是可控制“亮”、“熄”的象素。通过接插部分引入的信号电压，被加到相应的X、Y电极上，就能在显示平面上构成各种文字和图形。

② 交流型等离子体显示板的工作原理

交流型等离子体显示板的工作原理如图4·30所示。

当工作时，要在等离子显示板的全部X、Y电极间加上维持电压Vs，它是基频为10kHz至20kHz的方波，其幅值应低于气体的着火引燃电压V_f，但应能维持着火状态。当要引燃某一象素单元时，就在该象素的X、Y电极维持电压Vs上，再加上一个书写脉冲电压Vw，使Vs和Vw电压之和大于着火电压V_f，该象素单元就引燃发光。

在工作中，由于放电产生的电子和正离子积累到电极的保护层上，形成与外加书写脉冲电压反向的壁电压，见图（b）。当壁电压等于外加电压Vw时，两者抵消，放电熄灭。但当维持电压Vs到下半周时，与壁电压方向一致，两者叠加后又使该

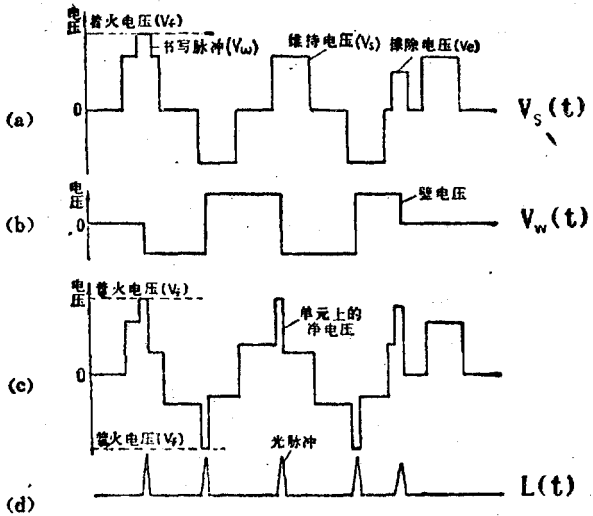


图 4·30 交流型等离子显示板的工作原理

象素放电发光，产生一个光脉冲。随后又开始形成与该瞬时电压方向相反的壁电压，又使放电熄灭。如此循环，便可产生一串光脉冲，使该象素单元发光。见图（d）。由此可见，只要有一个书写脉冲使某一象素发光后，不需反复加上同一书写脉冲，仍可由维持电压来保持发光。

如要使该象素熄灭，只需在适当位置加上一个擦除电压脉冲Ve，见图（a），使它同壁电压形成微弱放电而中和。由于消去了壁电压，虽有维持电压Vs也不能再继续放电，遂使象素单元熄灭。

这种仅加一次书写或擦除脉冲，就可使显示器发光或熄灭，尔后自行维持的特性，称为显示器的存储或记忆性能，这是等离子体显示器件的一个重要优点。

与液晶显示相比，等离子体显示的亮度高，对比度性能好，基本上不受使用环境光照的影响，尤其适合在黑暗环境中工作；与电子束管显示相比，等离子体显示的体积小，重量轻，视角大，功耗低，而且，显示面积可以做得比较大。目前，2048×2048象素的交流型等离子体显示板和1024×512象素的直流非存储型等离子体显示板都已有生产，在各种计算机终端和文字图形显示中得到了较广泛的应用。等离子显示的缺点是工作电压较高（约200伏），不能与集成电路直接相连，驱动电路比较复杂，成本较高。

（3）场致发光显示器件

场致发光是指某种物质（荧光粉）加上电场后能发光的物理现象。

场致发光可分成两类：一类是结型场致发光。这是一种在电场作用下少数载流子向P—N结的另一方注入，同多数载流子复合而产生的发光现象。这类器件称为注入式发光二极管（简称LED）；另一类是本征式场致发光。是用电场直接激励电子，使电子同空穴复合而产生的发光现象。这类发光现象常分成以下三种：将荧光粉掺入电介质中，加交流电压使其发光的称为粉末场致发光；工作时加直流电压，电极直接同荧光粉接触，不需用电介质的称为直流场致发光；采用薄膜技术制成的发光显示板称为薄膜式场致发光。图4·31是一种本征式粉末场致发光显示板的结构示意图。它是由玻璃衬底、条状电极（X向、Y向电极）、发光层和非线性层等部分组成。发光层是将选定的发光粉（多数属于硫化锌系列的材料）掺在介电常数很大的绝缘材料中，用有机溶剂调成稀糊状，再涂在条状的X向和Y向电极表面，形成厚度约为60微米的发光层。当在X、

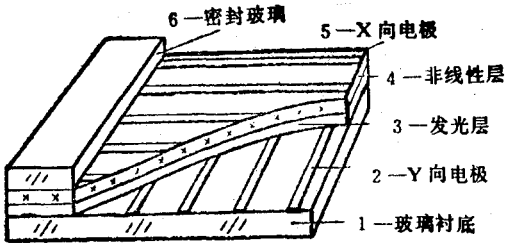


图 4·31　场致发光显示板结构示意图

Y电极上施加一定的电压值时，电极交叉点的象素就会受到电场激励而发光，形成文字和图形的显示。

实现场致发光显示的关键是如何解决交叉效应问题。所谓交叉效应，就是当某一象素发光时，与该象素相邻的其余 8 个象素也会受激发光的现象。这些象素的亮度虽然较低，但会严重影响显示图形的对比度，甚至造成错误的显示输出。产生这种现象的根本原因是发光粉的亮度随着外加电压逐渐升高而上升，没有一定的阀值。为了减少交叉效应的不良影响，在场致发光显示板的发光粉层上，再涂上一层厚度约为10微米的非线性电阻层，见图4·31中的（4），可以使发光粉具有一定的开关特性：当外加电压较小时，电阻层的阻值很大，接近于断开状态；一旦电压上升到一定值（阀值）时，电阻层的阻值变小，就使某一象素发光。而周围相邻的象素因为激励电压低于阀值而不发光。

场致发光显示器件的发光效率高，寿命长，工作电压比较低，可以做成大面积的文字图形显示器，具有广阔的应用前景。

4. 汉字显示器的基本构成和刷新方式

在以上两小节中，我们讨论了可用于汉字和图形显示的各种显示器件的基本原理和性能特点，简要介绍了常见的阴极射线管显示器和液晶显示板的扫描和驱动方法。这是实现汉字图形显示的第一步，解决显示器件在外来电压作用下发光（或改变光学特性）问题。本小节将讨论实现汉字图形显示的第二个问题：用什么方式实现对显示器件的控制和刷新，以进行正确的显示。

（1）汉字显示器的基本构成

一个完整的字符图形显示器，除了要有显示器件及其扫描驱动电路以外，还应有控制器、缓冲存储器和字符发生器等基本组成部件。图4.32所示为一个图形式阴极射线管（CRT）显示器的构成示意图。

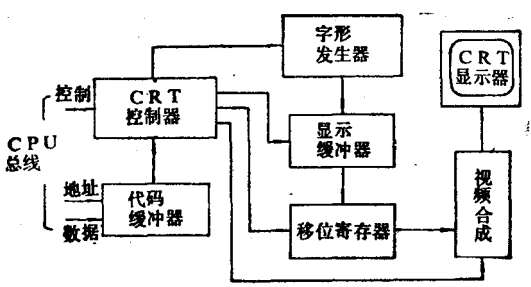


图 4·32 图形式CRT显示器构成示意图

显示控制器是显示器与主机CPU之间的接口电路，是整个显示器的控制中心。它根据CPU的命令，从代码缓冲器中

取出需在屏幕上显示的字形代码，再从字形发生器读出相应的字形点阵数据，并把这些数据存入显示缓冲器中。

显示缓冲器存放着与显示屏幕象素一一对应的数据信息。输出时，按照控制器提供的移位信号时序，通过移位寄存器变成串行信息流，再提供给视频信号合成电路。

视频信号合成电路把字形点阵信号、消隐信号、同步信号合成为视频信号（或称视频图象信号），加到CRT显示器件的调制极上，用以控制屏幕上的每一象素的“亮”与“暗”，便能显示出所需的字符图形。

在汉字显示器的各个组成部件中，控制器具有十分重要的作用。它除了提供一般显示器所需的各种同步扫描信号和消隐控制信号外，还要根据显示器件的不同特性、屏幕大小及所显示字形的点阵结构，选择各种不同的时序控制信号，这些都是保证汉字显示器能够显示出一幅稳定清晰和符合格式要求的字符图形所必不可少的。由于大规模集成电路技术的进展，现在已有多种单片集成的显示控制器可供选用。例如Intel8275和MC6845可编程CRT控制器，使用时只需附加很少的硬件便能与多种光栅扫描体制的显示器相连。显示器的每帧行数、每行字符数以及水平扫描线数等参数，都可通过编程由用户自行选定，使用比较方便。

（2）汉字显示器的刷新方式

如前所述，除了少数具有记忆性能的显示器件外，对大多数显示器来讲，屏幕上每个象素的“亮”、“暗”状态是不能持续的，必须有外加信号电压的反复扫描，才能保持其发光或不发光的状态。也就是说，必须循环反复地把视频信号加到显示器调制极上去，即需要进行显示器的屏幕刷新。

为了保持屏显图形的清晰和稳定，在进行屏幕刷新时应满足以下两点要求：一是必须使扫描信号与视频信号保持同步，即每次加上去的视频信号应与上次信号的时序严格保持一致，否则会使显示字形产生移动现象以致模糊不清。在汉字显示器中，各种同步信号都是由同一个时钟信号源逐级分频获取的，所以这一点要求比较容易达到。二是必须使屏幕刷新速度达到每秒30次以上，这样才能使显示的字形图象稳定，不会感到有闪烁的现象。但是刷新速度过快，使扫描频率（帧频和行频）增高，会造成控制电路复杂化。前面讲到的隔行扫描方法，就是一种在保证显示图象稳定、不产生闪烁感觉的前提下，尽量降低扫描频率的措施。

在实现以上两点要求的前提下，显示器的刷新过程就是反复读出显示缓冲器存储内容的过程，因此，这个显示缓冲存储器又称为刷新存储器。按照刷新存储器内信息的不同形式，汉字显示器的刷新控制方式可分为图形式和字符式两大类。

采用图形式刷新控制方式的汉字显示器（参看图4·32），显示缓冲器中存储的是待显示字形（图象）的整屏点阵信息，缓冲存储器中的每一位信息，直接对应显示屏上的一个象素单元。显示输出时，CPU 通过控制器首先把整屏显示数据全部读入显示缓冲存储器中。在显示刷新过程中，把整个屏幕的字形看作是一幅点阵图形，不断地从显示缓冲器中读取数据进行显示输出。由于在刷新过程中不再与字形发生器打交道，因此对字形发生器的读取速度要求不高，比较适合使用大容量磁盘系统作为字形发生器的情况。尤其在要求文字和图形混合显示的场合，采用图形式刷新方式比较方便。

图形式刷新控制方式有两个缺点：一是刷新存储器的容量比较大。对有384×640点阵的屏幕而言，要求存储容量大于30

K字节。二是在进行屏幕编辑过程中，画面的更换速度比较慢。例如，要增删一个汉字，或者要求汉字行滚动时，往往需要把刷新存储器的内容进行移位处理，处理的数据量有时是很大的，比较费时。因此，在需要经常更换显示内容的场合，图形式刷新方式使用起来不够灵活方便。

采用字符式控制方式的汉字显示器，显示缓冲器中存储的是待显示字形的代码信息。即需要屏幕显示的每个汉字或字符，在缓冲存储器中存放着它们相应的内部码。这种显示器的结构框图如图4·33所示。

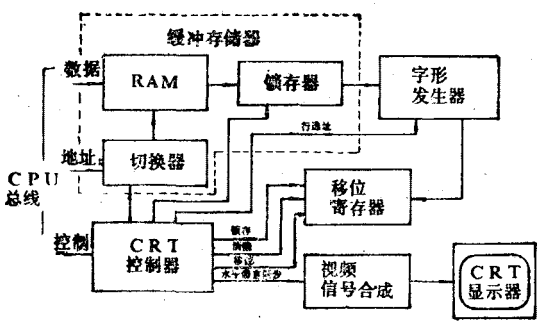


图 4·33　字符式CRT显示器的结构框图

每次刷新显示输出时，CRT控制器根据CPU的控制命令，按地址读出显示字形代码，经锁存器向字形发生器读出相应的字形点阵信息，再经移位寄存器变成串行的视频信号送往显示刷新。

字符式刷新控制方式的缓冲存储器所需的存储容量比较小。一个汉字的内部码一般为2字节，屏显字数为640个汉字的显示器，显示缓冲器的存储量不大于2K字节。由于缓冲存储器存放的是与屏幕显示位置相对应的汉字代码，因此，当进

行屏幕编辑时，对存储器的管理操作比较容易实现。

字符式刷新方式的缺点是对字形发生器的读取速度要求比较高。就以显示一幅640个汉字计算，若以每秒50次的速度进行刷新，则每秒要访问字形发生器640×50＝32000次；每字以72字节计算，字形发生器的数据读出速度为每秒32000×72＝2304000字节。显然，必须采用高速的ROM式字形发生器方能满足要求。在西文显示器中，由于字符数量不多，字形发生器结构很简单，读取速度能完全满足要求，一般都采用字符式刷新控制方式。在需用汉字量很大的系统中，如果使用磁盘存储器作为字形发生器，则需在内存中再开辟一个屏幕缓冲存储区，才能满足屏幕快速刷新的要求。

4.4 汉字打印机

汉字打印机是由打印执行机构及其控制、驱动电路组成的汉字输出设备。它能把机器内部的汉字及字符代码信息变换成书面的文字材料。打印输出是汉字终端和汉字微型机的一种重要输出方式。与显示器输出相比，打印机输出速度较慢，机械结构复杂，但是，由于书面文字材料具有可读性、可保存性等特点，使汉字打印机成为汉字终端设备的一个重要组成部分。而且随着打印设备制造工艺技术的提高，打印设备成本的下降，打印机的应用范围必将更加广泛和普及。

汉字打印机与西文的字符式打印机相比，基本的工作原理及结构大体相同，但在使用性能上各有特点。本节拟先讨论打印机的分类和特点；然后讨论适用于打印汉字的点阵式打印机的使用要求和具体性能；最后介绍几种常用打印机的基本工作原理。

1. 打印机的分类和特点

如前所述，打印机的基本功能是把文字代码信号变换成字形，并以所需的质量及格式打印在纸上。实现以上基本功能的打印设备种类很多，按照它们不同的打印输出方式来分，有行式打印机和串行打印机两大类。

（1）行式打印机

行式打印机简称行打机，它是计算机系统的外围设备之一，作为计算结果的高速打印输出设备。它能每次同时打印出一行字符或数据，打印速度每分钟可达几百行到几千行。按打印装置的不同结构分，行式打印机有鼓式、链式、列车式、钢带式和点阵式等多种。行式打印机的打印输出速度很快；但结构复杂，价格昂贵，在汉字终端和通用的汉字微型机系统中很少使用。

（2）串行打印机

串行打印机是采用串行方式打印输出的设备。所谓串行方式就是一个字符接一个字符地打印。因此打印输出速度比行式打印机慢，一般每秒打印几个到几十个字符。但是由于串行打印机的结构比较简单，价格比较便宜，已广泛应用于作为小型和微型计算机系统的外围输出设备以及各类终端的文字打印输出设备。

以不同的成字方式分，串行打印机有字模式和点阵式两大类。

① 字模式打印机

字模式打印机是把需打印输出的各种字形及符号刻制（或压铸）在特制的活字模上，再将这些字模装在一个能够灵活转

动（或移动）的支架及小车上。打印机工作时，根据不同的字符代码，控制字模支架的转动角度及小车的移动位置，使该字符代码相应的活字模与印字锤对准，通过印字锤迅速击向色带和纸，便可印出字形来。按照活字模支架的不同构造，字模式打印机有菊花瓣式、球式、杆式、轮式等多种、它们的共同特点是通过一次击打动作，便可印出一个字符。由于字模是整体刻铸而成的，印出的字形质量高，因此字模式打印机在字符打印输出的场合应用比较广泛，但是在汉字打印输出方面已不多见。如前所述，早期试制的汉字电传机，其打印机构也曾采用字模式，就是把几千个汉字刻成字模做在一个圆柱形的字鼓上，终因体积太大，运转不灵，未能达到实用水平。

② 点阵式打印机

点阵式打印机是应用点阵组字的原理来实现打印输出各种字符图形的设备。众所周知，用一定密度的点阵，可以组成各种形状的字符和图形。我们已在4·2节中介绍了用点阵数据描述字形，以实现汉字存储的方法。按照同样的原理，人们可以用打印一定密度点阵的办法，实现汉字及各种图形字符的打印输出。这种以点阵打印方式输出字形的设备就是点阵式打印机。

点阵式打印机的最大特点是记录的灵活性大。只要打印的点阵数足够多，可以印出任何一种形状复杂的文字或图形。而且在不改变任何硬设备的条件下，只需通过软件设计，便能改变输出文字的形状和大小，使用十分方便。因此不但在汉字打印输出、而且在字符和图形打印输出方面，都已得到广泛的应用。点阵式打印机的缺点是对打印头的质量要求高，制造工艺比较复杂。因为点阵式打印机的输出字形质量，除了字形点阵密度的因素外，与打印头的精度关系很大。打印的“点

子”愈细，点与点间的距离愈小，则打出的字形质量愈好。因此在要求打印字形质量较高的场合，需选用打印头精度为每英寸180点或300点的点阵式打印机。

按照不同的打印原理来分，点阵式打印机又可分为击打式和非击打式两大类。

击打式的点阵打印机又称为针打机。它是用电磁力推动一列钢针的后部，使钢针前端击向色带，在纸面上印出点阵字形的一种印字设备。针打机的主要部件是装有一列钢针的打印头。钢针数愈多，打印头的结构就愈复杂。打印西文字符和数字的打印头针数一般为7～9根，故打印头的结构比较简单，价格便宜；打印汉字的打印头通常有16～24根甚至更多的钢针，其结构比较复杂，价格较贵。

击打式打印机工作时噪音较大，耗电量较多；但只需用普通打印纸，成本低，而且打印的文本质量较好，能够长期保存。随着打印头制造工艺的日臻完善和打印字形质量的提高，针打式打印机在汉字输出打印方面的应用已日益普遍。

非击打式印字机是通过极细的传感头（打印头）把热、电等能量加到有特殊化学性质的纸或色带上，由于化学作用使纸面象素产生色差实现印字输出的设备。由于它不是“打”出字形的，故严格地讲应称为“印字机”或“印刷机”。但大多数仍习惯称为打印机。

非击打式印字机的种类很多，按照它们的作用原理来分，有热敏式、热转印式、静电式、激光式和液晶式等多种。喷墨式印字机是用极小的墨点喷射在纸面上印出字形的，也可归于非击打式印字机一类。热敏式和热转印式印字机是通过发热点使纸上产生黑色象素印字的，它们的印字头结构非常简单、体积极小，印字后不需对纸进行特殊处理，故常作为许多小型终

端的打印输出设备，但由于元件的热惯性时间所限，印字速度较慢。静电式和激光式印字机是通过静电或激光束调制和扫描，形成文字的潜象，再经过显影、定影等过程印出字形。它们都有极高的印字速度，但设备复杂，价格昂贵，多用在较大规模的计算机输出设备和照排系统中，在汉字终端和通用的汉字微机系统中应用较少。不过，近年来已研制出性能优良的台式激光印字机和液晶印字机，价格为针式打印机的四倍左右。由于印字质量高，使用方便灵活，深受用户欢迎。随着工艺和技术的进步，各种新型优质的打印机必将得到更广泛的应用。

现将几种常用的点阵式打印机的主要原理、性能和特点归纳如下，见表4-10。

我国是印刷技术的发源地。我国古代发明的活字印刷术对世界文化曾作出过巨大贡献。但是，汉字打字机和计算机汉字打印设备的普遍应用还是近十来年的事。其原因，除了现代工业和科技发展落后外，汉字本身的技术处理比较困难也是一个重要因素。现在，在吸收国外先进技术的基础上，终于解决了用机器打印汉字的难题。小型轻便的汉字点阵式打印机已经得到普遍应用，而且在机器的体积、重量和成本等方面都已达到或接近同类西文设备的水平，这必将对我国的经济、科技和文化的发展起有力的推动作用。

2. 点阵式汉字打印机的使用要求和主要性能

点阵式打印机作为文字输出设备，不论是打印西文字符或者汉字，其基本结构和工作原理都是大体相同的。但由于汉字在数量上比西文字母多，字形结构上比西文字母复杂。因此，作为汉字打印输出的点阵式打印机，还有一些不同于西文打印机的特点和使用要求，现分述如下：

表4·10　　常用点阵式打印机性能对照表

打印方式	打印(记录)原理	记录电压	记录纸	印字速度字/秒	清晰度线/毫米	记录可保存性	主要特点
针打式	电磁铁推动钢针击打色带和纸面	低	普通	10～100	3～11	好	工艺成熟，可靠性好，价格便宜；体积大、功耗大、噪音大。
热敏式	用感热印字头在特殊纸上加热显现墨点	低	特殊	10～300	4～20	差	体积小、功耗低、噪音小；特殊纸价格贵。
热转印式	用感热印字头加热，把色带墨点蒸发到纸上	低	普通	10～300	4～20	好	体积大，功耗低、噪音小，用普通纸；特殊色带较贵。
喷墨式	用极细喷嘴使墨水粒子化，经电场偏转后打到纸面记录	高	普通	30～120	4～12	好	体积小，功耗低，噪音小；特殊墨水和喷嘴的制造工艺复杂
电灼式	用放电破坏纪录纸表面白色层，使呈现黑点	低	特殊	10～300	对比度差	好	与热敏式相同，但还有臭味散出，有害健康

（1）汉字打印机的精度与打印字形质量

我们在4·2节中讨论了汉字字形质量与点阵数的关系，即增加字形点阵数目，可以表现出汉字的某些特性和风格，提高字形质量。但是，要做到有高质量的字形输出，还必须提高点

阵打印机的精度才能解决。

点阵打印机的精度，又称为分辨率，是指在单位长度内能打印的点（线）数，通常以每英寸（或毫米）若干点或线来表示。点阵打印机的精度愈高，它能打印的点子愈密集，能分辨的字形笔画结构愈精细，打印的字形质量自然就愈美观。例如，要用24×24点阵打印出5号字形（5号字的字身尺寸为10.5点，约3.7毫米），必须采用精度不低于180线/英寸的打印机才有可能。如果打印机的精度不够，即使有了24×24点阵字形数据，也无法在规定的字形尺寸内把它打印出来。表4·11是不

表4·11　　不同精度打印机采用字形点阵值建议表

铅字号	铅字字身尺寸（点）	折算毫米数		180线/英寸	240线/英寸	300线/英寸	372线/英寸	480线/英寸	540线/英寸
		字身	字面	取值	取值	取值	取值	取值	取值
超特大号	72.000	25.30	23.28		224	272	384	448	512
特大号	63.000	22.14	20.37		192	240	340	384	448
特中号	56.000	19.68	18.11		172	214	302	340	398
特初号	48.000	16.87	15.52		144	184	256	302	340
特　号	45.000	15.82	14.55		136	172	240	272	320
小特号	42.000	14.76	13.58		128	160	224	256	300
初　号	36.000	12.65	11.64		108	136	192	224	256
小初号	30.000	10.54	9.70		92	112	160	184	214
一　号	28.000	9.84	9.05	64	86	108	150	172	200
小一号	24.000	8.44	7.76	56	76	92	128	150	172
二　号	21.000	7.38	6.79	48	64	86	112	128	150
小二号	18.000	6.33	5.82	40	56	72	96	112	128
三　号	15.750	5.54	5.10	36	48	64	86	96	112
四　号	14.000	4.92	4.53	32	40	56	76	86	96
小四号	12.000	4.22	3.86	28	36	48	64	76	86
五　号	10.500	3.69	3.39	24	32	40	56	64	76
小五号	9.000	3.16	2.90	20	28	36	48	56	64
六　号	7.875	2.77	2.55	16	24	32	40	48	56
七　号	6.000	2.10	1.93		16	24	32	36	40

说明：表中计算值与取值单位均为点，是字面尺寸，不是外框尺寸。'点'是常用字型单位，1点=0.0353mm。

同精度的点阵打印机容许采用的字形点阵数建议表。

由表4·11可知，精度高的点阵打印机，能在相同的字形尺寸内容纳更多的点阵数目，打印高质量的点阵字形。如果打印机的精度不够，采用相同的点阵数目时就必须扩大字形尺寸，这样会使笔画（特别是斜线笔画）的不连续性和不光滑性突出起来，使人感到字形“发虚”，降低了字形输出的质量，而且长时间阅读极易产生眼睛疲劳的感觉。

随着汉字终端的迅速推广应用，人们对汉字字形质量的要求愈来愈高，甚至希望有接近于字模排版印字的清晰度。为了能满足这些要求，就要用提高点阵式打印机精度的办法来解决，这是生产、使用汉字点阵式打印机必须认真考虑的问题。

（2）汉字打印机的成字方式

用点阵打印机打印汉字，是以点阵位图方式组成各种字形的。它的基本方式是以一排直列的钢针同时横向扫描打印成字。根据打印钢针数和打印字形点阵数的不同情况，汉字打印机有整列打印、拼接打印和交叉打印三种成字方式。

① 整列打印成字方式

整列打印成字方式是用一排直列的钢针一次打印出字形点阵的一个纵列，再通过打印头的横向移动，逐列打印出一个完整的字形。用整列打印成字方式时，点阵打印机的针数必须等于或大于点阵字形的行数；当打印头的横移次数与点阵字形的列数相等时，就完成了一个汉字的打印输出。

例如，用24针打印机打印一个24×24点阵的汉字，需要打印头驱动钢针打印及横向位移24次。用同样打印机打印一个15×16点阵的汉字，则需要打印头驱动一部分钢针（不超过16根），并横向移动15次。

整列成字打印方式是汉字点阵打印的主要方式。用这种方式打印的字形点阵结构稳定，质量好，输出速度快，一般每秒达30～40个汉字，甚至更多些。但是，整列打印方式的输出字形变化较少，当字形的纵列点阵超过打印机针数时，就只能用拼接成字方式打印输出。整列成字打印方式对打印头的制造精度要求较高，必须使钢针之间保持极小的间距，才不会使打印出的字形发虚。此外，当你需要用8针或9针的字符式点阵式打印机打印汉字时，就不可能使用整列成字方式，而只能用拼接打印成字方式了。

② 拼接打印成字方式

拼接打印成字方式是用一列钢针一次打印字形点阵纵列的一部分，通过打印头逐步向右横移，打出字形上部的n行点阵（n是打印头的针数）；在打印纸上移和打印头左移以后，再打出字形的下部n行点阵。重复以上过程，直到拼接打印成整个点阵字形为止。

例如，用9针打印机打印24×24点阵字形，必须分三次打印拼接成字。第一次，打印头向右横移(24+s)×M步*，打出一整行汉字字形的上部9行点阵。然后打印纸上移9行点阵的距离，并使打印头左移；第二次，在打印头右移(24+s)×M步的过程中，紧接着打出字形的中部9行点阵。再使打印纸上移、打印头左移；第三次，用打印头的上面6根钢针，紧接着打出字形的下部6行点阵。经过三次打印过程，就可拼接打印出完整的24×24点阵字形。

通过拼接打印方式可以用针数较少的点阵打印机打印较大字号和较多点阵数值的汉字，也可以实现图形打印输出。由于

---

* s是字与字之间的点距数，M是一行汉字数。

针数较少的针打机价格比较便宜，（例如9针打印机的价格是16针打印机价格的$\frac{1}{2}$～$\frac{1}{3}$，或是24针打印机价格的$\frac{1}{5}$～$\frac{1}{6}$），因此，用拼接打印方式打印汉字目前仍有不少应用，特别是在大号字形和某些图形的打印输出方面，仍是不可缺少的一种打印方式。

拼接打印成字方式的缺点是输出字形质量不稳定，常常产生拼接字形的脱节或错位现象，影响字形美观，甚至造成差错。事实上，要使一个字形点阵拼接达到紧密无间的程度，就必须对打印头的平移速动和输纸行距（即打印纸的上移运动）的误差控制相当严格，要排除在打印过程中各种外来干扰因素的影响，这是一个相当困难的任务。此外，由于拼接过程中的打印头往返运动，使打印输出速度明显减慢。因此，在经济条件可能的情况下，正常字体的汉字打印输出，应尽量不用这种拼接成字方式。

③ 交叉打印成字方式

交叉打印成字方式，是用装有两排直列钢针的打印头交叉打印同一列字形点阵，然后再通过打印头的横向移动，逐列打印出一个完整的字形。图4·34是交叉打印成字方式示意图。

如图(a)所示，是用1列7针打印头整列打印“E、X”两个字符，每个字符打印头右移5次。图(b)是用2列6针打印头交叉打印“X、E、g”三个字符的情形：当打印头移到打印位置时，右边1列6根钢针先打印出字形点阵第一列的偶数点子；把打印头右移一步，用左边一列6根钢针打印出字形点阵第1列的奇数点子；再用右列钢针打印字形点阵第2列的偶数

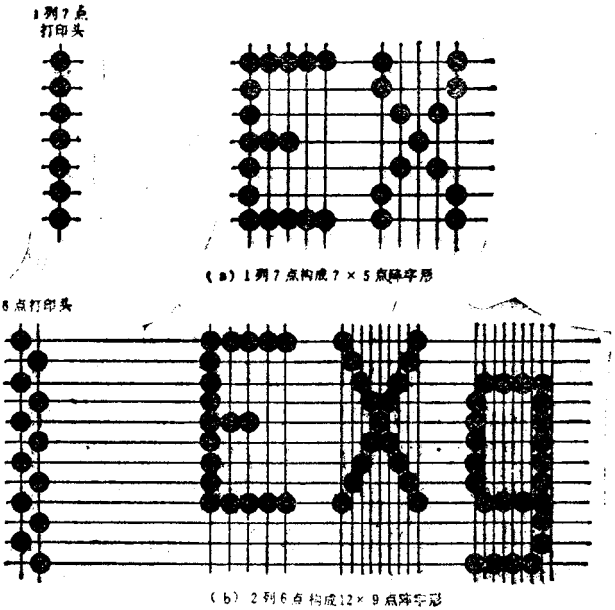


图 4·34 交叉打印成字方式示意图

点子；……这样逐点移动打印头，交叉打印偶数点和奇数点，就可实现字形打印输出。

由图可见，用交叉打印成字方式，可以使同一直列中的奇数点与偶数点略带重叠，再通过控制打印头移动间距的办法使同一横行中的点子也略带重叠，就能使点阵字形呈现较好的笔画连续性，克服字形发虚的毛病，有效地提高点阵式打印机的输出字形质量。因此，绝大多数点阵式打印机都采用这种交叉打印成字方法。

与其它打印成字方式一样，交叉打印成字方式对点阵打印

机的字车移动间距和输纸精度的控制要求也比较严格，否则就会出现竖笔弯曲或笔画断开等现象，影响打印字形的质量。

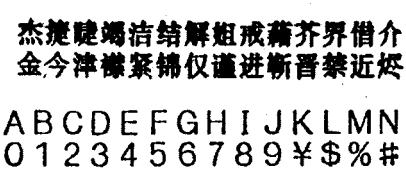
（3）汉字的字体字型和文件格式处理

汉字打印机打印输出的文件资料，除了要求有良好的字形质量以外，还应有对字体、字型以及文件格式的处理能力，这也是衡量汉字打印机质量的一个重要内容。

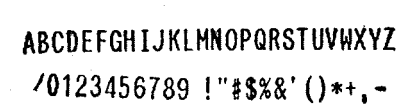
通常情况下，一份文件资料总是有标题、正文、注释等不同部分组成，需要用不同的字形尺寸或字型来打印。有时在同一份文件中，对某些重要内容需要用黑体字或仿宋体字打印，以示醒目。在科技文献资料中则需要用到各种外文字母、公式符号等不同字符和图形。因此，对汉字打印机提出各种不同的使用要求。

如前所述，汉字的常用字体有宋体、仿宋体、楷体、黑体四种，一般通用的汉字终端和汉字微机的打印设备都装有GB5199、GB5007或GB6345国标汉字点阵字形发生器，可以提供标准宋体字形的点阵数据。在档次较高的汉字打印设备中，作为正式打印文件使用时，还装有其它三种字体的字形发生器，可供打印不同字体字形之用。在只装有一套标准点阵字形发生器的汉字终端和汉字微机系统中，虽然只能打印一种宋体字，但是可以通过打印程序的设计来放大或加宽字形，使之成为竖粗横细或等线体的倍宽字型，如图4·35所示。

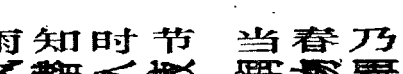
图（a）是一种标准字型，字的高度略大于字宽，呈微长形，是文件正文中用得最多的字形，可以按照字形发生器输出的点阵数据直接打印。图（c）的倍宽字型是在标准字形的基础上，使字形高度的点阵数不变，字宽方向的点阵数加倍而打印出来的。与标准字形相比，高度不变，字宽放大一倍，可以



(a) 标准字型（全角）



(b) 半角字型



(c) 倍宽字型（横写、竖写）

图 4·35 常用打印字型

作为标题字型。如果在加宽水平点阵数的同时，把竖笔点阵重复打印一遍，就成为竖粗横细的倍宽字型，可使字形结构更显得美观些。

当文件资料中需打印字母、数字和符号时，可以选择标准字型(又称全角字型)，如图(a)所示；或者半角字型，如图(b)所示。标准字符的字体宽高都与汉字一样，半角字符的字体高度不变，宽度的点阵数为标准汉字点阵的$\frac{1}{2}$。这样可以使汉字与字符混合打印输出时的文件格式保持整齐。例如，在一个汉

字行中若出现奇数个半角字符时，可使打印头自动移过相当于一个半角字符的间格，便能保持各行间的字形上下始终对齐。这种全角和半角字符的点阵数据，都可从字形发生器直接取得。

字形打印输出时的排列方式是汉字打印机特有的格式处理问题。西文拼音字母都是从左至右横排的，现代汉字也是用横排格式，即以从左至右顺序排成“行”来阅读的，但是有不少场合仍然沿用传统的直排格式，即以从上至下顺序排成“列”来进行阅读。为了适应这一特殊要求，汉字打印机应有横排和直排两种打印输出格式，如图4·35（c）所示，上边是横排格式，下边是直排格式。实现两种不同格式打印的方法是在读出字形发生器的点阵数据以后，在打印缓冲存储器里将数据排列方式旋转90°。

汉字打印机一行能打印的最大字数，取决于打印头在一行中平移的最大点距数和汉字标准字型的点阵数。例如某型号汉字打印机的打印头能从左至右移动960个点距，则一行能打印的最多汉字数为：15×16点阵标准字型60个，或者24×24点阵标准字型40个。若行中有非标准字型，则以一个倍宽字型相当2个标准字型计算。汉字打印机的其它各种格式控制功能，将在下一章介绍国标GB5261—85《文字和符号成形设备用的增补控制功能》时进行讨论。

3. 汉字打印机的基本原理及构成

汉字点阵式打印机通常由缓冲存储器、打印控制器、驱动电路和打印执行机构(打印头)四个基本部分构成，性能完备的汉字打印机还包括专用的字形发生器或打印汉卡。它们的构成框图如图4·36所示。

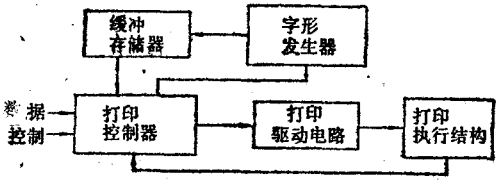


图 4·36　汉字点阵式打印机构成框图

（1）打印缓冲存储器

打印缓冲存储器是一组专用的随机存取存储器(RAM)，

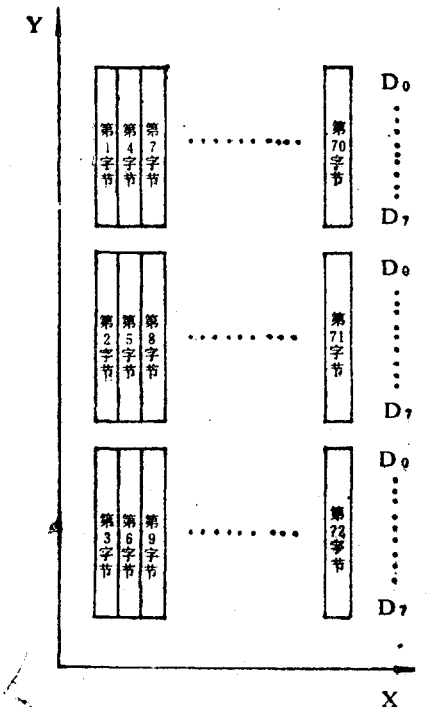


图 4·37　24×24点阵字形数据直排格式

它的作用是把字形发生器送来的点阵数据按照打印机所需要的格式存放起来，然后供打印控制器将当前一行的打印数据送入打印驱动器，驱动打印执行机构（打印头）打出字形。

打印缓冲器的存储容量通常以一行汉字的点阵数据计算。例如一行可打印60个15×16点阵汉字的打印机，它的缓冲存储容量可为32×60=1920字节。即相当于满行时一行汉字字形数据的信息量。当一行汉字打印完毕后，主机CPU根据打印控制器送来的信号，再从字形发生器取出下一行汉字的字形数据送入缓冲存储器，开始下一行汉字的打印。

由于针打机的打印执行机构是一排直列的钢针，一根钢针的驱动信号对应于点阵数据字节中的一“位”（即比特），因此，要求缓冲存储器中的点阵数据字节也是直排的，即每一字节是按低位到高位的顺序从上至下排列。一个24×24点阵字形的直排数据格式如图4·37所示。读者可对比图4·25的字节横排格式，两者正好相差90°。如果在缓冲存储器中以字节横排格式的次序存放点阵数据，则打印的汉字字形将会旋转90°，就是文本直排格式输出了。

（2）打印控制器

打印控制器是主机与打印执行机构相连的接口部件。它接收主机送来的各种打印控制信息和汉字代码信息，把它们变换成打印驱动电路的控制信号，去驱动打印执行机构完成打印操作。一种打印控制器的控制信号及其时序如图4·38所示。

图中DB0～DB7是数据总线，DATA表示总线上出现了准备打印的数据。当过了20ns即总线上数据稳定后，主机在$\overline{\text{STB}}$（选通）线上发出一个＞0.5μs的负脉冲信号，即选通信号，将数据总线上的数据送入打印控制器；在选通信号开始后

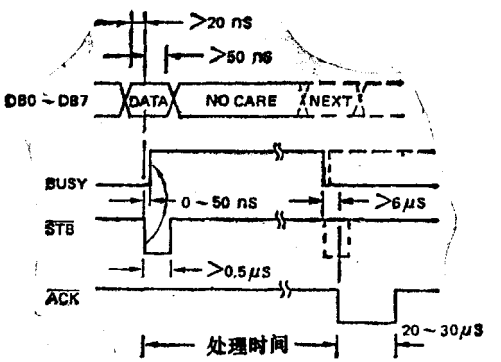


图 4·38 一种打印控制器的控制信号及其时序

0～50ns内，打印控制器在BUSY（忙）线上向主机发回一个正脉冲信号，表示打印机已经开始打印，正忙。此信号一直持续到打印过程结束，在BUSY信号有效期内，主机不再向打印控制器送数据。BUSY信号结束后，如果需打印下一个数据（NEXT），主机再用$\overline{\text{STB}}$通知打印控制器继续第2个数据的打印过程；如果没有数据了，打印控制器向主机发回$\overline{\text{ACK}}$（认可）信号，表示打印机已经完成数据打印操作，主机可以据此信号向打印控制器发送新的数据或控制命令。

除了上述的基本控制信号外，汉字打印机还需一些专用和通用的格式控制信号如：

- 字形的扩大或缩小；
- 横向打印或竖向打印；
- 中西文混合打印格式；
- 加下线及背景网点式样；
- 水平及垂直制表参数设置。

以上控制信号不是由专用的控制线传送，而是由打印控制器根据主机送来的控制命令，再转换成打印执行机构的相应驱动信号。

在打印控制器与打印执行机构之间的控制信息也有多种，基本的控制信息有：

- 回车（CR），把打印头返回本行行首位置；
- 换行（LF），把打印纸上移一行距离；
- 新行（NL），同时实现回车和换行；
- 换页（PF），把打印纸上移一页（66行）距离；
- 间隔（SP），把打印头水平推进16个点距；
- 汉字间隔（KSP），把打印头水平推进24个点距。

除了以上基本控制信息外，打印控制器对打印执行机构的控制信息还有：行距、字距和行宽的设定，打印速度选择，宽页对半，自动分列，版面禁止规则处理等等。

由此可见，随着点阵式打印机功能的不断完善，对打印控制器的要求愈来愈高。现在已经普遍采用微处理器作为打印控制器的核心，并把缓冲存储器、字形发生器都纳入打印控制器统一管理。这样不但能容易地实现各种复杂的打印控制功能，而且可以减轻主机CPU的负担，提高汉字终端的工作效率。这种打印控制器的硬件结构框图如图4·39所示。

由图可见，打印控制器CPU通过内部总线连接2—3个并行接口、固化了控制程序的监控ROM、打印数据缓冲器RAM和字形发生器。并行接口的作用是：一个用于与主机交换数据和信息；另一个（或2个）用于与内部打印驱动器连接，向打印执行机构传送各种数据和命令。

监控ROM中的控制程序将控制命令转换成打印执行机构所需的操作信号。它包括两部分内容：一部分是字形发生器和

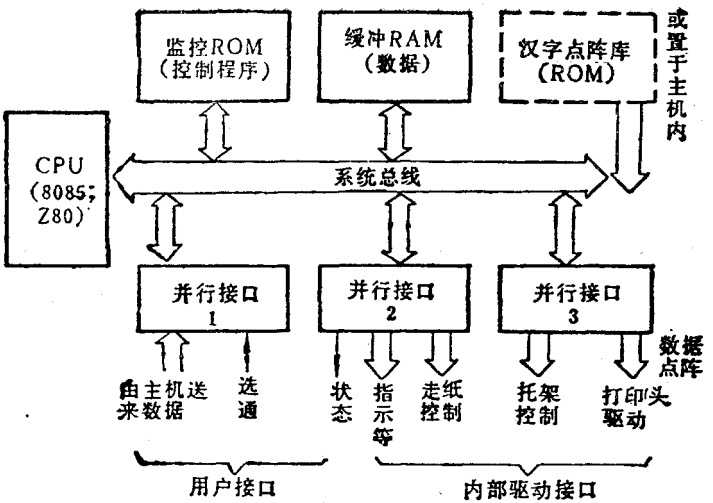


图 4·39 针式打印控制器硬件结构框图

缓冲存储器的管理程序，主要用来把主机送入的汉字代码变换成字形发生器输出的点阵数据，然后以打印执行机构所需的格式放入缓冲存储器。另一部分是打印操作控制程序，主要用来完成打印文字的操作控制。它利用定时器保持对打印头当前位置的跟踪，从缓冲存储器取出一列点阵信息送入打印驱动器，以驱动打印钢针动作。打印执行机构的其它各种格式控制功能和操作功能，则以相应的服务子程序来实现。

把字形发生器设置在打印控制器内的好处是显而易见的。这样一来，主机便可把待打印的汉字代码（每字 2 字节）送入打印控制器，不再传送几十字节的字形点阵数据了，从而可以节省主机CPU的操作时间，提高汉字终端的效率，而且还有利于实现打印控制程序的兼容。目前市上出售的各种打印汉

卡，它的主要部件就是一组字形发生器及固化的汉字打印软件。设计精良的打印汉卡，不仅使普通的点阵打印机能打印汉字，而且使它的功能得到很大的扩展。随着固化字库的普及应用，配置专用字形发生器的汉字打印机将会更加普遍。

（3）驱动电路

驱动电路是打印执行机构的电气控制部分，它把打印控制器送来的各种信号加以放大和整形，去驱动相应的执行机构，实现字形打印输出和完成各项操作。其主要功能是：

·向打印头的各个打印电磁铁提供驱动信号；

·产生字车移动步进电机和输纸步进电机的各相励磁驱动信号；

·产生控制打印机各种状态所需的驱动信号。

驱动电路的主要功能框图如图4·40所示。

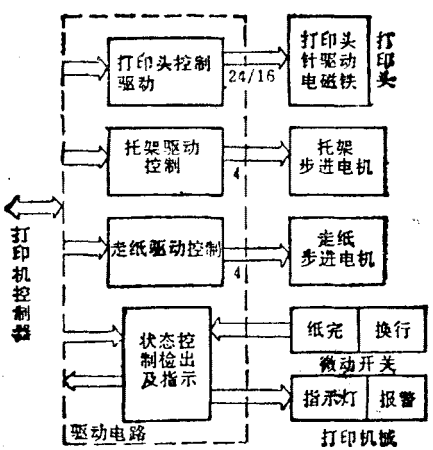


图 4·40 针式打印机驱动电路主要功能框图

图中的打印头控制驱动电路，是按照缓冲存储器输出的一列点阵数据，去控制相应的逻辑门，启动复合功率放大管输出励磁电流，从而驱动电磁铁拍打钢针完成打印动作。例如，从打印缓冲器取出当前打印的一列数据值是3B78H（十六进制值），它的二进制值为0011101101111000，用这16个位值控制16个逻辑门，位值为“1”时使逻辑门启通，位值为“0”时使逻辑门关闭，于是可以使第3、4、5、7、8、10、11、12、13共九个电磁铁得到励磁电流，推动9根钢针动作，印出1列16个点中相应的9个墨点。

托架驱动控制电路提供两种字车驱动信号：一种是正向打印时的字车驱动信号，即在打完前一列点阵后，字车带动打印头右移一个点的间距，随后接着打印第二列点阵。另一种是字车反向运动（回车）的驱动信号。当打印头到达一行终点位置，或者根据打印格式需要另起一行时，字车带动打印头左移，回到一行的起始位置上。字车驱动信号加到三相或四相步进电机上，通过机械传动装置实现字车的运动。当字车正向运动时，一般是等速等距移动，移动过程中字车所在的当前位置应通过传感装置反馈到打印控制器，以确定取出下一步打印的数据；当字车反向运动时，应使驱动电流脉冲频率升高，字车先作加速运动；当字车到达减速微动开关位置时，步进脉冲按减速的逻辑降低频率；当到达一行的起始位置时，触动行首微动开关，使字车运动停止。

走纸驱动控制电路的工作过程，与字车驱动电路基本相同。

点阵式打印机对上述两个驱动电路的信号脉冲精度及稳定性的要求比较高。尤其在采用拼接印字或交叉印字方式时，字车托架移动和走纸距离的误差，将使印字质量严重下降。

状态控制检出及指示电路向打印机控制器提供打印机的状态信息，如打印纸是否用完，换行时的走纸距离（行距）应多宽等等。同时把机器状态的报警及指示信号送到有关的单元电路中去。

（4）打印执行机构

打印执行机构是汉字打印机的执行部件，它是由打印头、托架（字车）传动装置、色带传动装置和打印纸传动装置等部分构成。打印执行机构的结构如图4·41所示。

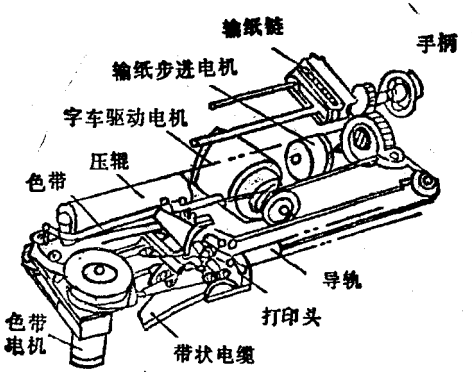


图 4·41　打印执行机构示意图

打印头是点阵式打印机的关键部件，它的基本功能是把一个个电流脉冲变换成打印纸上的一个个墨点。不同结构的打印头具有完全不同的“打点”方法，因而成为不同类型的点阵式打印机。如钢针式打印机（针打机）、热敏式打印机、喷墨式打印机等等。我们将在下一小节具体介绍几种常用打印头的结构和工作原理。图4·41所示是一种钢针式打印机。

托架（字车）传动装置是由字车驱动步进电机、托架、导轨等部件构成。打印头被安装在托架上，能够沿导轨左右移

动，好象一架小车，故又称字车。字车驱动电机通常是一台步进电机或直流伺服电机，在打印控制信号的“指挥”下，将字车传送到预定位置，用打印头在纸上打出墨点。字车传动装置中还包括能检测字车位置的传感器。打印控制器根据传感器提供的位置参数决定提供哪一组打印数据，并通过带状电缆传送到打印头上。

打印纸传动装置由输纸步进电机、输纸链、压辊、手柄等部件构成。它的作用是带动打字纸在垂直方向上下移动，使打印头能在纸上一行一行地打字，并根据需要保持适当的行间或页间距离。输纸步进电机是传动装置的动力源，它按照打印控制器规定的时间和距离传动打印纸。传动的方式有两种，一种为齿轮传动，用于带边孔的打印纸；另一种为摩擦传动，用于不带边孔的普通打印纸。当需要人工上纸或退纸时，则用手柄操作。

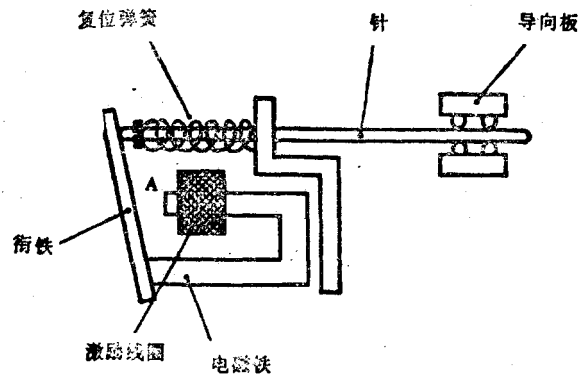
色带传动装置由色带电机、色带盘等部件构成。它带动色带在打印头与纸面之间移动，使打印针每次击打在色带的不同部位，保证墨点颜色清晰。使用多色色带的打印机，色带传动装置除使色带平移外，还能根据控制命令使色带上升到不同高度的位置上，可供打印机实现多种颜色的打印输出。热敏式和喷墨式打印机不需色带及传动机构，其打印执行机构的结构比较简单。

4. 常用点阵打印头的工作原理

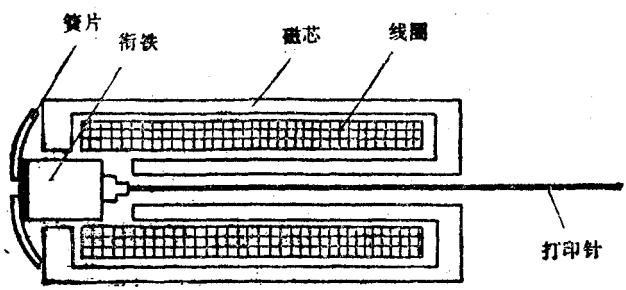
（1）钢针击打式打印头

钢针击打式打印头简称针打头，是目前应用最广的一种点阵打印头。它的基本构成部件是排成直列的一组硬质合金钢针，其一端（针头）穿过宝石轴承导孔靠近色带及打印纸，另

一端（针尾）受电磁铁及复位弹簧的控制。常见的针打头结构有两种，一种是拍打式，另一种是螺管式，它们的内部结构示意图如图4·42所示。



（a）拍打式



（b）螺管式

图 4·42　两种针打头的内部结构示意图

图（a）是拍打式针打头的内部结构示意图，它采用一个马蹄形电磁铁作为钢针打印的动力源。平时，在复位弹簧的

作用下使钢针左移，针头离开纸和色带，针尾左推衔铁，使它与电磁铁之间保持一个距离A。当有打印脉冲电流输入电磁铁的激励线圈时，电磁铁产生瞬间集中的强磁场吸引衔铁向右运动，克服复位弹簧的阻力把打印针推向色带，并在纸面上打出一个墨点。为了使打印墨点保持准确位置，在靠近针头处设有导向板，引导钢针的动作。导向板内还装有宝石轴承，以保证钢针在高速频繁的往复运动时有足够长的使用寿命。

图（b）是螺管式针打头的内部结构示意图，它采用一个螺管形的电磁铁作为钢针打印的动力源。衔铁在螺管内部，并与钢针尾端固定在一起。平时，衔铁左端的簧片使钢针复位，针头离开色带和纸。当输入打印脉冲电流使电磁铁励磁时，磁芯被磁化后吸引衔铁，推动打印针向右运动，击打色带，并在纸上印出墨点。

由多个螺管式打印针构成的打印头外形结构如图4·43所示。

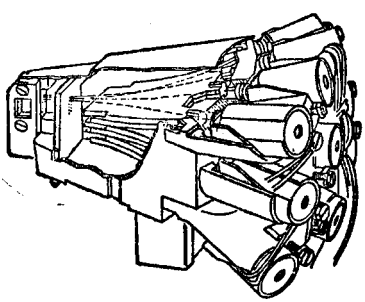


图 4·43 针打头外形结构

由图可见，16个打印螺管电磁铁的后部组成圆形结构，前部16根打印针通过导向板上的小孔被分成2列，采用交叉打印成字方式，可以得到较好的字形质量输出。

除了打印针的排列方式外，打印针的质量、针头直径、针间距离和针的动作特性等因素，都与打印点阵字形质量及打印头的寿命密切相关。打印针的材料通常采用质地坚韧耐磨的合金，应保证在正常使用条件下击打动作1.5亿次以上。针头的直径是打印机分辨率（精度）的决定因素之一，在目前的加工工艺条件下约为直径0.2mm。打印机分辨率愈高，在字号一定时能容纳的点阵数就愈多，打印的字形质量就愈好。但限于条件，把针头做得更细还有困难。为了提高打印输出速度，对打印针的动态特性，包括电磁铁驱动特性、钢针的动距大小、针与支架轴承的配合等因素，应进行周密的综合设计。

（2）喷墨式印字头

喷墨式印字头简称喷墨头，是喷墨式印字机的关键部件。按照喷射墨点的不同方式，喷墨头有连续喷墨式和冲击喷墨式两种。

① 连续喷墨式

连续喷墨式印字的基本原理是：用一个极细的喷嘴连续地喷射墨水微粒，在墨粒的运动过程中用电场（或电荷）控制其偏转方向，使它喷着到纸面的预定位点上，形成字形输出。图4·44是一种电荷控制型连续喷墨印字头工作原理。

如图所示，墨水容器中的墨水经微型泵加压至墨水喷管。置于管内的压电振子被115kHz晶振电压激励振动，使墨水呈谐振状态，形成微粒经墨水喷管头的喷嘴射出。喷出的墨水微粒通过带电电极时，由于加在电极上的视频信号电压受到汉字字形发生器的调制，使墨粒带上了与字形信息相应的电荷。接着，带电墨粒通过有高电压的垂直方向偏转板产生向上偏转运动，因而到达纸面预定位置形成墨点。不带电荷的墨粒继续保

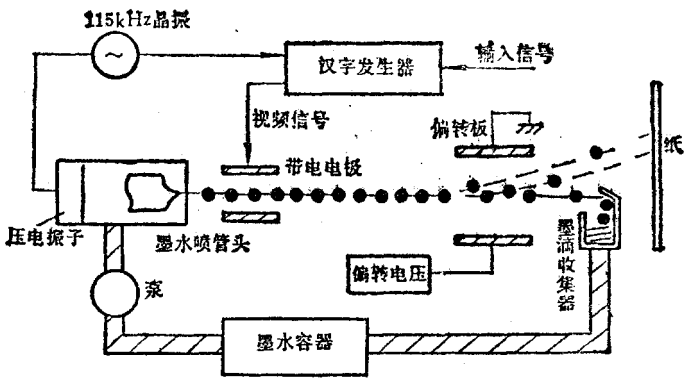


图 4·44 电荷控制型连续喷墨印字工作原理

持水平运动方向，由墨滴收集器回收到墨水容器中。偏转板上的偏转电压变化一个周期，就形成汉字点阵字形的一个纵列墨点。再通过托架带动喷墨印字头的水平横向移动，就可以形成一个完整的点阵字形。也可以用再加一对水平偏转板的办法，使墨粒产生横向偏转，印出一个完整字形后，再移动印字头打印下一个汉字。

② 冲击喷墨式

冲击喷墨式印字的基本原理是：用具有多个墨水喷嘴的喷墨印字头，通过类似针打头的作用原理，以控制每个喷嘴是不是喷射墨粒的办法来形成字形点阵。冲击喷墨式印字头的外形结构如图4·45所示。

由图可见，一组靠近纸辊表面的墨水喷嘴，也象钢针一样排成两列，并经过喷水通道与墨水瓶连通。控制墨粒是否喷射的部件是套在墨水通道上的压电陶瓷管。这是一种压电换能器

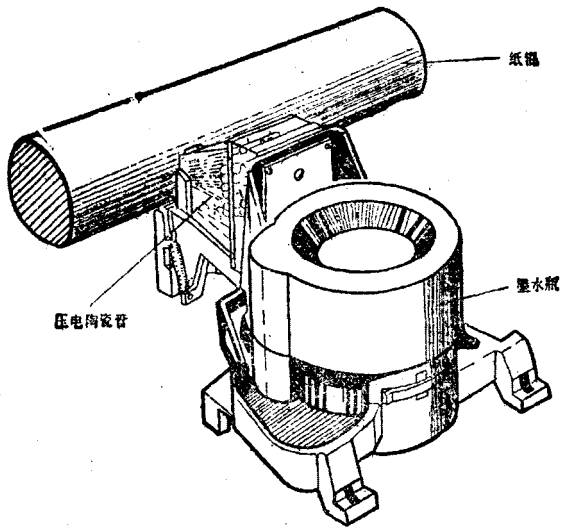


图 4·45 冲击式喷墨印字头外形结构

件，受字形发生器送来的数字脉冲电压控制。当需要某一喷嘴喷墨时，就在该喷嘴墨水通道的压电陶瓷管上施加一脉冲电压，使陶瓷管产生瞬时形变挤压墨水通道，墨粒就从喷嘴射出。

与连续喷墨式相比，冲击喷墨式印字头的结构比较复杂，需用多个喷嘴。而喷嘴的加工工艺要求很高，因为它的直径极细，内孔直径只有50微米左右。但是冲击式喷墨印字机的控制过程比较简单，印字机结构与针式打印机类似，价格也已不相上下，而且喷墨印字速度快，字形质量好，印字时无噪音，是一种极有应用前景的汉字打印方式。目前的主要问题是，墨水喷嘴的加工工艺技术难度大，特殊墨水的配方需改进，一般墨水在空气中易生沉淀，堵塞喷嘴，影响正常喷墨印字。

（3）热敏印字头

热敏印字头是热敏式印字机和热转印式印字机的关键部件。它的结构极为简单，实质上就是一个通电后能产生若干发热点的集成电阻块。按照不同的制作方式，有薄膜电阻式、厚膜电阻式和半导体扩散电阻式三种。其中薄膜电阻式可以做成具有精细图形的印字头，而且它的热惯性较小，因而应用较广。

热敏式印字头的工作原理是：用代表字形点阵信息的数字脉冲电流使印字头上的发热点瞬时发热，这些发热点也类似打印针排成1列或2列，当某些发热点瞬时发热时，与色带或打印纸相触，在纸面上印出1列中的若干墨点。通过印字头在托架上的横向水平移动，便能印出点阵字形。

热敏打印纸是热敏式印字机的特殊用纸，它是用无色的有机染料和酚类化合物作为发色剂，将它们研磨成微粒喷涂在纸基上加工而成的。热敏印字头与热敏纸如图4·46所示。

当热敏印字头贴近热敏纸时，印字头上的发热点使热敏纸受热变色，变色的反应速度与发色剂及发色色素的含量有关。一般情况下，当印字头发热单元电阻为100欧姆、脉冲电流持续时间为4毫秒时，便可使受热点发色部分呈现墨点。

由于热敏印字头重量体积极小，且可省去色带及其传动机构，使热敏式印字机的结构简单，体积小、重量轻、耗电省，工作时无钢针打字的噪音干扰，适合于移动条件和需保持环境安静的场合应用。热敏印字机的缺点是需消耗价格较贵的热敏打印纸。为了克服这一缺点，发展了一种不用热敏打印纸的热转印式印字机。

热转印式印字机的结构及原理与热敏式印字机完全相同，

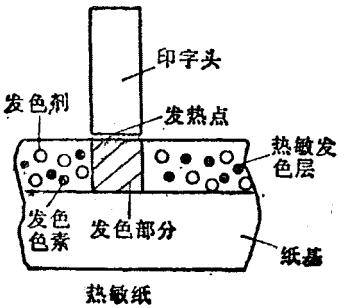


图 4·46　热敏印字头与热敏纸

唯一的区别是用热敏印字头通过热转印色带在普通打印纸上印出点阵字形。它的工作原理是：当热敏印字头的发热点受脉冲电流加热时，与它相接触的热转印色带上受热点的墨色蒸发转印到打印纸上，形成墨点。这种热转印色带涂有一种特制的感热墨色涂料，大多只能一次性使用，且价格较贵。如果热转印色带用完，把普通打印纸换成热敏打印纸，可把热转印式印字机作为热敏式印字机使用。

五、汉字代码及汉字通信

在第一章中我们讨论了汉字终端各种代码的性质、作用以及它们之间的相互转换。在第三、四两章中又具体讨论了两种汉字代码——汉字输入码和汉字字形点阵数据码的结构特点及其产生和存储方法。本章将着重讨论汉字终端和汉字微型计算机的内部码、传输交换码、各种控制功能码，并介绍有关汉字通信的基本知识。

5·1 汉字内部码和传输交换码

汉字内部码是汉字终端或汉字计算机系统在进行运算、存储等信息处理过程中应用的一种汉字代码。通常，在机器内部处理过程中需要对汉字信息进行各种频繁的操作；并且要求汉字代码与计算机系统原有的西文代码充分兼容。因此，必须正确设计和选用汉字内部码。

众所周知，人们主要通过字形来识别和区分汉字。汉字的字形代码是以数据字节串的形式存储在机内，这就是字形点阵数据码。用这种代码可以方便地控制显示器或点阵打印机，以实现汉字的字形输出。但是，字形点阵数据码太长，不能用来作为内部码。如前所述，用来表示一个汉字字形的点阵数据至少需要32个字节，即16×16=256位。如果在机器内部操作处理过程中也直接采用字形点阵数据码，这将会大量耗费机器资源（即机器的操作时间和内存空间），是非常不合算的。

就以西文字符来讲，虽然一个字符的字形点阵数据只需七、八个字节，比汉字字形数据量要少得多，但是在计算机处理字符信息时，也是需要用字符的内部码进行操作的。只有当进行字符的字形输出时，才将字符的内部码转换成字形点阵数据码。

为了能够在两个或两个以上的汉字终端或汉字计算机之间进行信息交换，还需要有适合于线路传输的交换码。这是同内部码既有联系又有区别的另一种汉字代码。内部码只是在本机内部使用，设计考虑的角度是如何使机内处理操作简便，并与系统原有的代码兼容；交换码涉及两个以上的不同机器或系统，应有统一的标准。如果两个系统相距较远，还必须考虑到便于线路传输及通信的要求，这是内部码与交换码的不同之处。但是，由于这两种代码之间关系十分密切，相互转换非常频繁，因此，应尽可能使它们之间具有简单的一一对应关系，以提高机器操作效率。在不少西文设备和系统中，往往直接采用交换码作为内部码，就是这个缘故。但是，对于汉字系统来说，情况要复杂得多。因为现有的汉字系统几乎都是在西文系统的基础上进行二次开发而成的，而汉字交换码也是在西文交换码的基础上扩展编成的。如果直接用汉字交换码作为汉字内部码，就不能做到汉字系统与西文系统的兼容，因此必须进行汉字内部码的设计。为了能具体说明这个问题，我们还必须从西文系统的内部码谈起，再介绍信息处理交换用的七位标准代码和两种汉字交换码，最后重点介绍汉字内部码设计的基本要求及几种内部码方案实例。

1. 西文系统的内部码

西文系统的内部码是表示西文字母、数字及常用符号（统

称字符）以便于计算机处理的数据代码。人们把需用的字符排成一张表格，用一组有序的二进制数对应编号，就成为一张代码表。用不同的字符排列次序和不同的编号方法，可以得到各种不同的代码表。由于常用的字符（加上基本控制功能字符）总数不超过256个，故用不超过8位二进制数(即1个字节）便可实现字符编码。

根据不同的编码规则，可以有不同的代码表设计。显然，用于两个系统之间进行信息交换的代码表设计必须一致，否则就不能交流信息。早在1963年，美国标准学会制订了包括128个字符的7位代码表，作为美国信息交换标准代码，即America Standard Code for Information Interchange，简称ASCII（“阿斯克”）码。以后，国际标准化组织（ISO）和国际电报电话咨询委员会（CCITT）也颁布了内容大体相同的代码表，简称国际5号码。我国国家标准局在1980年颁布实施的国家标准《信息处理交换用的七位编码字符集》，即GB1988—80，编码规定与以上几个标准都大体相同。所谓“大体相同”，就是在代码表中，除货币符号（编码为24）表示的形状各异外*，其余的字符编码都完全一样。因此在一般情况下，当提到ASCII码、国际5号码或GB1988码时，都是指的同一个代码表，即七位标准代码表，或称为标准的七位信息交换码。

国内通用的七位标准代码表如表5·1所示。

如前所述，七位标准代码表最初是作为信息交换用的字符代码设计的。随后，它作为计算机信息处理用的字符代码，成为大多数系统的西文内部码。这样一来，使西文系统的信息处

* 货币单位的符号在ASCII码中为$，在国际5号码中为¤，在GB1988码中为¥。

表 5·1　国内通用的七位标准代码表（GB1988码）

b_4 b_3 b_2 b_1	b_7 0	0	0	0	1	1	1	1
	b_6 0	0	1	1	0	0	1	1
	b_5 0	1	0	1	0	1	0	1
	0	1	2	3	4	5	6	7
0 0 0 0	0		SP	0	@	P	`	p
0 0 0 1	1		!	1	A	Q	a	q
0 0 1 0	2		"	2	B	R	b	r
0 0 1 1	3		#	3	C	S	c	s
0 1 0 0	4		¥	4	D	T	d	t
0 1 0 1	5		%	5	E	U	e	u
0 1 1 0	6		&	6	F	V	f	v
0 1 1 1	7		'	7	G	W	g	w
1 0 0 0	8		(	8	H	X	h	x
1 0 0 1	9		)	9	I	Y	i	y
1 0 1 0	10	LF	SUB	*	:	J	Z	j	z
1 0 1 1	11			+	;	K	[	k	{
1 1 0 0	12			,	<	L	\	l	|
1 1 0 1	13	CR		-	=	M	]	m	}
1 1 1 0	14			.	>	N	^	n	~
1 1 1 1	15			/	?	O	_	o	DEL

理过程与交换过程的码制相同，避免了相互转换的麻烦。但是，七位代码表的容量小，七位信息与计算机的字长不一致，作为内部码使用时不够方便和合理。于是，就产生了其它各种不同方案的西文内部码，其中著名的是IBM公司计算机系统采

用的EBCDIC码。

EBCDIC是Extended Binary-Coded-Decimal Interchange Code的简写。EBCDIC码的代码表如表5·2所示。

表 5·2　　EBCDIC代码表

b_7 b_6 b_5 b_4	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
0 0 0 0					SP	&										0
0 0 0 1									a	j			A	J		1
0 0 1 0									b	k	s		B	K	S	2
0 0 1 1									c	l	t		C	L	T	3
0 1 0 0									d	m	u		D	M	U	4
0 1 0 1									e	n	v		E	N	V	5
0 1 1 0									f	o	w		F	O	W	6
0 1 1 1									g	p	x		G	P	X	7
1 0 0 0									h	q	y		H	Q	Y	8
1 0 0 1									i	r	z		I	R	Z	9
1 0 1 0					¢	!										
1 0 1 1						$	.	#								
1 1 0 0					<	*	%	@								
1 1 0 1					(	)	-	'								
1 1 1 0					+	;	>	=								
1 1 1 1						¬	?	"								

从表5·2可知，由于采用了8个二进制位（8位）编码，EBCDIC代码表的容量比ASCII代码表容量扩大一倍，可以表示256个字符组成的字符集。代码表的编码规律性也比较好，它把一个字节的8位分成两部分，前半部分为列号，后半部分为行号。代码表中的任一字符，可以用分别代表行号和列号的两个16进制数表示。如大写字母“A”代码的二进制值为11000001，即16进制数0C1H；小写字母“a”代码的二进制

值为10000001，即16进制数81H；依此类推。表中的空白项00～3F为控制字符代码，其它可以作为扩充的字符编码使用。

当使用EBCDIC码作为机器内部码的设备与其它设备进行信息交换时，必须先把EBCDIC码转换成标准的七位信息代码，即通用的信息交换码。

由于七位信息标准代码的应用十分广泛，它既是西文系统的交换码和内部码，又是构成汉字信息交换码的基础，因此有必要对这个代码表作进一步讨论。

2. 七位标准代码表

如上所述，七位标准代码首先是作为西文信息交换码而制订的，代码表容量为128个字符，具体内容见表5·1。由表可见，任一个字符分别用代表行号的4个二进制值和代表列号的3个二进制值来表示。前者称为低位码（b_4～b_1），后者称为高位码（b_7～b_5）。

例如大写字母“A”，所在行号为0001，列号为100，也就是“A”代码的二进制码值为1000001即16进制值41H；小写字母“d”的所在行号为0100，列号为110，“d”代码的二进制码值为1100100，即16进制值64H；依此类推。任一字符的七位代码结构如图5·1所示。

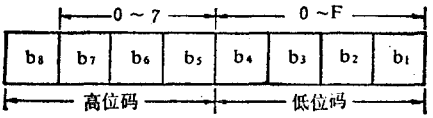


图 5·1 一个字符的七位代码结构

在计算机中，通常以8位二进制值组成一个字节。当用来表示一个字符的内部码时，这个字节的第8位（最高位b_8）是

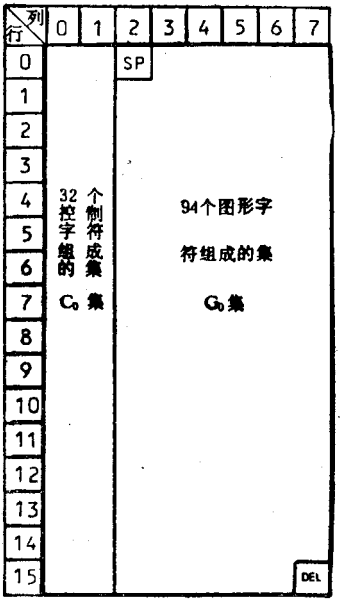


图 5·2　七位代码表的构成

空着不用的，通常把它置 0，即$b_8 = 0$。

从整体上看，一张由 8 列、16行构成的七位代码表可以划分成控制字符集和图形字符集两部分，如图5·2所示。

由图可见，高位码为000和001两列是控制字符集，或称C_0集，包括32个控制功能字符，它们的具体控制功能和含义将在5·2节中介绍；高位码为010到111的六列是图形字符集，或称G_0集，除了左上角的SP字符和右下角的DEL字符（其功能也在5·2节中介绍）外，共有94个图形字符。其中包括大、小写字母各26个，阿拉伯数字10个，以及常用数学符号、商业符号和标点符号32个。在这些图形符号中，除了代码24H所表示的

货币符号在各个具体标准中各不相同外（见第276页注＊），其余127个字符，不论是图形字符还是功能字符，在ASCII码、国际5号码和国标GB1988码中的符号、码值及含义都完全相同。

值得注意的是，当七位标准代码在作为内部码时，每个字节的最高位（b_8）是闲置不用的（通常令$b_8=0$）。换句话说，在$2^8=256$个可编码位中还有80～FF的128个码位可以使用，但目前尚无统一标准，由计算机系统的设计者自行安排。我们将在关于汉字内部码的讨论中涉及这个问题。

3. 汉字交换码

汉字交换码是用于汉字信息系统或汉字终端之间交换信息的数据代码。众所周知，为了进行信息交换，交换码必须采取统一的格式。

由于历史的原因，我国国内用于汉字通信的代码目前有两种格式。一种是电报码，是以《标准电码本》规定的汉字代码；另一种是国标码，是以国标GB2312-80规定的汉字代码。

（1）汉字电报码

汉字电报码用于汉字电报通信，已有一百多年的历史。当国外使用莫尔斯电报机和电传打字机进行电报通信时，我国以四个数字代表一个汉字作为通信代码，对实现汉字电报通信发挥了重要作用。至今，电报码在邮电通信领域中仍然得到广泛的应用。

汉字电报码是按照《标准电码本》的规定，用0～9中的任意4个数字组成一个汉字代码，可以表示一万个汉字或非汉字图形字符。图5·3所示，是《标准电码本》中的一页。

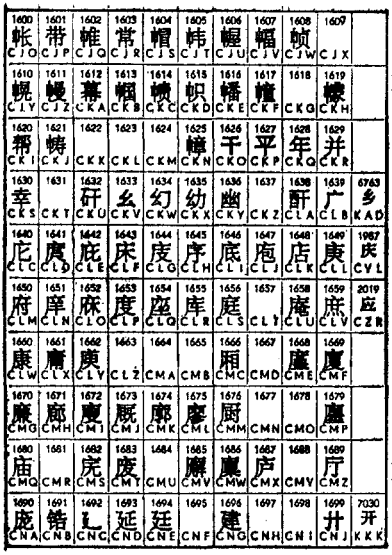


图 5·3 《标准电码本》中的一页

由图可见，汉字“平”的电报码为1627、“年”的电报码为1628、“府”的电报码为1650等。

在电报通信中，通常采用5单位编码，即用5位二进制值表示一个字符的代码。国际通用的5单位代码是国际电报电话咨询委员会的第2号代码表，简称为国际2号电码。我国除使用国际2号码外，还有一种专门为防止10个数字出现传输差错而设计的数字保护电码，简称为数字保护码。

当用4个数字一组的电报码进行汉字通信时，每个汉字的信息码位为5 × 4 = 20位（比特），为了保证每个字符代码的发收双方同步，还需在5个信息码位的前后，各加上1个起始位（1比特）和1个停止位（1.5比特），组成7.5比特的一个字节来传输，因此，一个汉字的实际传输信息量为7.5×4=30

位（比特）。汉字电报码的传输格式如图5·4所示。

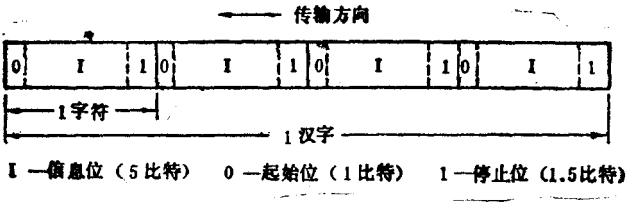


图 5·4 汉字电报码的传输格式

（2）汉字国标码

汉字国标码又称为汉字国标交换码，是国家标准《信息交换用汉字编码字符集—基本集》GB2312-80图形字符代码表的简称。这个代码表是在7位标准代码表（表5·1）的基础上扩展的双字节7位代码。它在1981年5月颁布实施，并已得到国际标准化组织（ISO）的认可，成为我国法定的信息交换用标准汉字代码。

汉字国标码目前已规定了信息交换用的通用汉字6763个、非汉字图形字符682个的代码。它以7位标准代码表中的G_0集（参见图5·2）为基集。也就是说，用码值在21H至7EH范围内的94个图形字符代码构成GB2312-80代码表中的纵座标和横座标。代码表中的每一个汉字和图形字符，都以它们所在位置的纵、横座标定位，用两个字节、每字节7位二进制值来表示。第1字节21H至7EH，以纵列从上到下排序；第2字节21H至7EH，以横列从左到右排序。整个代码表结构如图5·5所示。

由图可见，代码表纵向分成1～94个区，横向也分成1～94个区，可容纳94×94＝8836个汉字及非汉字图形字符。全表

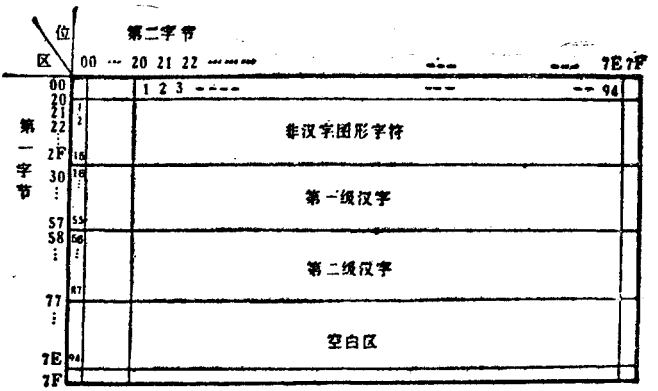


图 5·5　汉字国标码（GB2312-80）代码表结构

分成4个部分，以区号划分：第1区至15区为非汉字图形字符部分，第16区至55区为第一级汉字部分，第56区至第87区为第二级汉字部分。第88区至94区为空白区部分。上面三个部分中也有若干空白区位，留待今后增补之用。其中有些空白区位已经由国标汉字点阵芯片加以利用。

代码表的第一部分，目前安排了682个非汉字图形、字母、数字及各种常用符号，它们有：

- 一般符号202个，在1、3、9区；
- 序数号60个，在2区；
- 阿拉伯数字22个，在2、3区；
- 拉丁字母52个，在3区；
- 日文假名169个，在4、5区；
- 希腊字母48个，在6区；

• 俄文字母66个，在7区；

• 汉语拼音及注音符号63个，在8区。

代码表的第二部分，目前安排了常用汉字3755个，留有待增补的空位5个。由于常用字一般都知其读音，为便于查找，按照汉语拼音字母顺序排列。多字音取常用音，只排一次。同音字以声调排列；同音同调字以起笔笔形“横、竖、撇、点、折”为序排列。若起笔笔形相同，则依次取第2笔、第3笔笔形。

代码表的第三部分，目前安排了次常用字3008个。由于这些字大多不太熟悉其读音，因此以部首和笔画为序排列，部首划分与一般字典大致相同，略有改动合并。部首顺序以笔画数排列，同笔画数的以笔形“横、竖、撇、点、折”为序排列。变形部首排在正部首之后。同部首的字按除去部首以后的笔画数排列，同笔画数时也按起笔笔形排序。

图5·6是GB2312-80图形字符代码表（国标码）的一部分。在代码表中，任一汉字及非汉字图形字符均由它所在位置的区位号标识。区号与位号之间用连字符号相连，为该字的区位号。例如“京”字在代码表的30区、9位，它的区位号是30-09；“诘”字是二级字，在代码表的58区、21位，它的区位号是58-21。在代码表中的非汉字图形字符也是如此，例如数字“1”的区位号是03-17，其余依此类推。

我们应当注意国标码与区位号的区别。如前所述，国标码是以七位代码表GB1988中的两个字节表示一个汉字。由图5·6可见，标识纵座标的是第一字节，标识横座标的是第二字节，它们的码值范围是21～7E。例如，“京”字的国标码是3E29，“诘”字的国标码是5A35，阿拉伯数字“1”的国标码是2331等等，以上数字或字母都是16进制数。四位16进制数代表两个

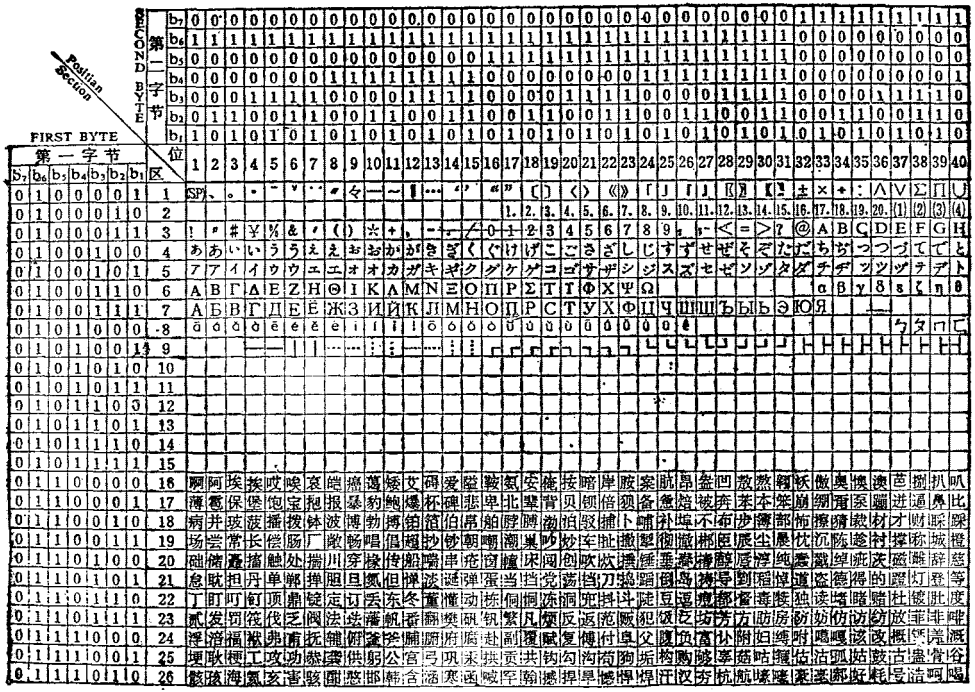


图 5·6　GB-2312-80图形字符代码表（部分）

字节。由于每个字节的最高位是0，所以一个汉字的有效信息位是14比特。

汉字国标码《基本集》（GB2312）中的6763个汉字的选择，是在我国文化部和前中国文字改革委员会于1965年联合发布的《印刷通用汉字字形表》6196个字的基础上，根据汉字信息处理需要增加了500多个科技、地名和姓名用字确定的。对于国内一般的汉字终端和汉字信息处理系统，《基本集》的汉字数量已可满足使用需要。但由于当时选字条件的限制，字表中尚有一百多个字的使用频度几乎为零，这是《基本集》的不足之处。

汉字的实际使用情况比较复杂，现用的和历代遗留下来的古字总数有五万多个。目前港、澳、台地区及某些使用汉字的其它国家还仍在使用繁体字和异体字。有些部门（如古籍整理和出版单位、城市户籍系统等）则对用字数量要求特别多，因此，仅有《基本集》是不够的，还必须制订若干个《辅助集》，才能满足上述地区和部门汉字信息处理的需要。

1986年9月，国家有关部门审定通过的国家标准《信息交换用汉字编码字符集——第二辅助集》和《信息交换用汉字编码字符集——第四辅助集》，这两个辅助集共收汉字14276个，其中第二辅助集7237字，第四辅助集7039字。

第二、四辅助集的汉字字形为通用的规范字形（简化字），不收繁体字和已经淘汰的异体字（未经整理的异体字收入第四辅助集）。由于这两集的汉字绝大多数是罕见字，一般不知其读音，为了便于检索和使用，均按照部首排列；同部首字，按笔画数排列。部首按1983年文改会和文化部出版局等单位联合制订的《统一汉字部首排检法草案》规定的201个部首排序。

根据我国目前规范字(简化字)与繁体字并用的实际情况，为了便于汉字信息的处理和交换，繁体字将与规范字分别对应编成代码集。规范字集的序号用偶数表示，即基本集、第二辅助集、第四辅助集；繁体字集的序号用奇数表示，即第一辅助集、第三辅助集、第五辅助集。原则上，只需将“规范字集”内的简化字改成相应的繁体字，就可成为“繁体字集”，以便于查阅使用。

综上所述，汉字国标码是作为汉字信息交换用的代码设计的，但是，在汉字信息处理时也经常需要用到它。

4. 汉字内部码的设计要求

如前所述，汉字内部码是计算机进行信息加工处理时用的汉字代码。在一个汉字信息处理系统中，汉字内部码的设计和选择，对系统的性能和效率关系很大。目前，对汉字内部码还没有统一规定，不同的汉字系统有各不相同的内部码设计方案。从现有的实践经验看，汉字内部码的设计和选择，应满足以下几点要求：应与国标码有简单明确的一一对应关系，以便于它们之间相互转换；应保证系统的中西文兼容；应有利于提高系统的效能。

（1）汉字内部码与国标码应便于相互转换

前已提到，汉字国标码是我国法定的汉字信息交换用的标准代码，而汉字内部码目前尚无统一规定。因此各个不同汉字终端和汉字微机系统要互相交换信息，就必须把内部码转换成国标码。目前大量生产的汉字字形发生器存储芯片，大多是以国标码作为读出地址的，就是说，要从字形发生器读取字形点阵数据，也必须把内部码转换成为国标码。总之，在汉字信息

处理和传输交换过程中，汉字内部码与汉字国标码之间的转换是十分频繁的。为了尽量提高系统效率和减少转换中可能发生的差错，应使内部码与国标码之间的转换规则尽量简单；并且两种代码一一对应，没有二义性。因此，汉字终端和汉字微机系统的内部码设计，都不是脱离国标码完全另搞一套，而是在国标码的基础上稍加改造，使两者保持简单明确的对应关系。

但是，内部码与国标码是不能完全等同的，如果为了免去转换的麻烦而使两者相同，则就会影响系统中西文兼容的要求。

（2）应保证汉字系统的中西文兼容

现有的汉字系统实际上都是在西文计算机系统的基础上开发扩充而来的，由于西文系统大多经过较长时间的开发和应用，具备丰富的软件资源，其中凝聚了人们开发这些资源的大量劳动和智慧。在西文系统基础上开发的汉字系统，当然应该尽可能保持已有资源，这对发展我国的信息处理事业是十分重要的，这就要求汉字系统必须中西文兼容。否则，一切都要从头做起，损失就太大了。

保证汉字系统实现中西文兼容的关键，在于应使汉字代码与西文代码在系统内部同时并存，互不冲突。也就是说，汉字内部码的设计，要使系统在处理信息时对这两种代码能够明确区分开。

但是，现用的汉字国标码（即GB2312码）是在西文字符代码（即ASCII码）的基础上制订的，它们之间相互紧密联系，不好区分。我们在国标GB2312-80的代码表（图5·6）中看到，汉字国标码是由两个ASCII码组成的。也就是说汉字国标码的码值与ASCII码的G_0集是重合的。因此如果直接把国标

码作为汉字内部码，则系统会把汉字代码与西文字符代码互相混淆起来。例如，当连续出现30和21(16进制数)两个字节时，系统将无法确定它们究竟是一个汉字“啊”的代码，还是两个字符“0”和“!”的代码。由于系统原有的大量软件都是用西文ASCII码写成的，新写入的汉字文件中也常有汉字与ASCII字符混合使用的情况，如果系统不能把汉字内部码与ASCII码区分开，势必造成系统的工作混乱，根本无法实现中西文兼容。

也许有人要问，在国标GB2312-80的码表中也有西文字符，这不是中西文兼容吗？的确如此，但不能称为中西文兼容。因为在GB2312的码表中，西文字符的编码与ASCII码表中的编码是完全不同的。仔细对比一下表5·1和图5·6可知，前者是用1个七位字节表示1个字符，后者是用2个七位字节表示1个字符。例如阿拉伯数字“5”的ASCII码值是35，而在GB2312-80中的码值是23、35两个字节。因此，如果把西文字符用GB2312的代码表示来作为系统的内部码，则系统中原有的ASCII码就失去了它们本来的含义，原有的系统软件当然无法运行，也就谈不上中西文信息的兼容了。

因此，必须有这样的内部码设计：当系统处理时能够识别和区分哪些是汉字代码，哪些是西文代码，这才能实现中西文兼容。总而言之，把国标码照搬过来作为内部码是不行的，脱离国标码另搞一套的内部码也是不可取的，必须根据具体条件作出恰当的选择。

（3）应有利于提高系统效能

汉字内部码在机内操作频繁，因此内部码设计与系统效能关系很大。为了提高汉字系统的效能，应尽量使码长最短；应

尽量采用等长码，并使码值连续有序。

众所周知，码长与编码字集大小密切相关。西文字符集较小，只需7位或8位码长便可实现编码。汉字常用的有几千字，内部码码长至少需要双字节的。双字节7位码最多可以表示$2^{14}=16384$个可区别码，双字节8位码最多可表示$2^{16}=65536$个可区别码。这对通常的汉字终端和汉字微机系统来讲，已经完全够用了。因此，可采用双字节的最短码长作为内部码。这不但能节省系统开销，提高效能；而且双字节码恰好与汉字的显示和打印特性互相配合，易于输出处理。但是，双字节码的兼容性较差。在一些较大的系统中往往不能满足使用要求，因此还需采用3字节或4字节码作为内部码。

采用码组等长的内部码（即等长码），能使机器在检索时有规律可循，提高操作速度。因为汉字内部码通常是以字节串的数据结构存放在机内的。当需要从字节串中查找某个汉字的内部码时，对于等长码就比较方便，只需指明该字在某页中的位置，便可推算出它的相应存放地址。例如，已知某字在本页第一行第n个字的位置，只需把页地址上加（2n－1）的地址数，就可把这个字检索出来。反之，如果内部码不是等长码，进行每一次检索操作都必须对字节串全部数据扫描一遍，那就要费事多了。

保持内部码码值连续有序，对提高检索查找操作的速度是十分必要的。例如，有100个内部码的字符串是无序排列的，若要从其中查找某一字符代码时，必须采取逐一比较的方法，最快第1次比较就查到了，最慢就要比较100次，平均比较次数为51次。如果这100个字符是按照递增或递减规律连续有序排列的，就可采用“折半查找法”查找，最快比较1次，最慢比较7次便可确定了，检索效率提高许多倍。而且字符串愈

长，效率提高愈显著。如果内部码不是连续有序，还会导致出现一个内部码对应多个国标码的二义性，使系统操作发生差错。

与任何复杂事物一样，汉字内部码设计的要求是互相制约的。在目前没有统一的内部码标准的情况下，设计者应从实际出发通盘考虑，采用比较适用本身情况的汉字内部码方案。

5. 汉字内部码方案介绍

如前所述，我国目前多数计算机系统和终端设备是在引进国外设备的基础上开发汉字功能的，由于原有系统性能的差异以及不同环境下的二次开发，汉字内部码的设计也就多种多样。概括地讲，可分为七位码方案和八位码方案两大类。现分别举其应用较广的几种扼要介绍如下：

（1）七位码方案

七位码方案是以《七位编码字符集》（GB1988码集）中94个图形字符的G_0集为基集，用两个或两个以上的7位字节构成汉字代码的方案。大体有以下四类。

① 用带标识符的国标码为内部码

这是目前应用较多的双字节或3字节七位码方案。

这类内部码方案采用GB2312-80代码表规定的汉字国标码作为内部码，但是要在出现国标码时（不论是出现一个或连续多个），在这个或这些代码组的前后必须加上特定的标识符，以便把汉字代码与西文字符代码区别开来。这个标识符选自七位代码表的G_0集，以不用或最少使用的一、二个ASCII图形字符来充当。

现通过讨论几个实例，说明用带标识符的国标码作为汉字

内部码的不同方法。

我们以“三角形ABC”一句为例子来说明。“三角形”3个汉字的国标码分别是：

三（487D） 角（3D47） 形（504E）

“ABC”3个字母的ASCII码分别是：

A（41） B（42） C（43）

括号里的码值都是十六进制数。为了使文字说明简明一些，在本段中都省去了后缀“H”或下标16。

方法一：假定选一般很少用的符号“～”作为标识符，它的ASCII码值为（7E），则“三角形ABC”的内部码表示为：

7E 487D 3D47 504E 7E 41 42 43

机器区分汉字代码与ASCII代码的方法是：首先必须随时记住标识符“7E”出现的次数，当“7E”奇数次出现时，随后出现的代码是汉字内部码；当“7E”偶数次出现时，随后出现的便是ASCII码。

方法二：假定选用符号“&”（26）作为汉字代码开始、ASCII码结束的标志；选用“$”（24）作为ASCII码开始、汉字代码结束的标志，则“三角形ABC”的内部码表示为：

26 487D 3D47 504E 24 41 42 43

采用方法二的优点是机器可以不必记住标识符的出现次数，但必须分别使用两个不同的标识符。

这两种内部码表示方法，汉字代码与ASCII码之间的界面清晰，汉字内部码的平均码长略大于两个字节，少于三个字节。但这些方法实现起来比较繁琐，系统必须随时对标识符进行检测和计数，标识符的丢失或意外增加，都会使整个文句产

生混乱。

方法三：为了减少因标识符的偶然变化引起的混淆，采用在每一个汉字代码前面都加一个标识符的办法，成为三字节的汉字内部码，即一个汉字内部码是由两个字节的国标码加上一个字节的标识符代码构成。假设选用“&”（26）为标识符，则“三角形ABC”的内部码表示为：

26 487D 26 3D47 26 504E 41 42 43

此法的优点是机器不需记忆标识符的出现次数，不易引起两种代码的混淆，系统的可靠性提高了，但使每个汉字代码的码长增加了，系统的效率随之有所降低。

以上介绍了三种用带标识符的国标码作为汉字内部码的方案，它们有一个共同的问题，就是当出现汉字代码与标识符代码相同的情况时，会发生混乱。因为国标码的两个字节都是由ASCII码表中（21）至（7E）范围内的全部可显示图形字符代码构成的。例如，上例中汉字“三角形”的国标码码值与ASCII码6个图形字符“H、}、—、G、P、N”的码值相同，其中有两个符号“}”和“—”如果也被用作标识符，机器就会无法识别。又如，在以“～”（7E）作为标识符的汉字串中，如果出现一个“溯”字（国标码3A7E），就有可能使机器产生计数错误而引起混淆。

为了克服带标识符的二字节或三字节七位码作内部码的缺点，进一步提高机器工作的可靠性，于是就产生了用四字节七位码作汉字内部码的各种方案。

② 四字节的七位码方案

四字节的七位码方案是用四个七位标准代码作为一个汉字的内部码。它的基本构成方法是在三字节内部码的基础上，将三个字节的第2字节和第3字节（即国标码的两个字节），经

过查表方法或简单算法变换成三个字节。这三个字节必须是字母或数字的ASCII码。在它们前面加上原有的标识符，就构成了四字节的内部码。四个字节的基本格式为：

标识符（符号）＋字母＋数字＋字母或数字

第1字节是标识符，可以选用ASCII码表中任意一个非字母数字的符号；第2字节至第4字节是从双字节国标码通过查表或计算变换取得的。其中第2字节一定为字母代码，第3字节一定为数字代码，第4字节则为字母或数字代码，但一定不是符号代码。图5·7是这种四字节内部码方案的字母数字混合编码表。

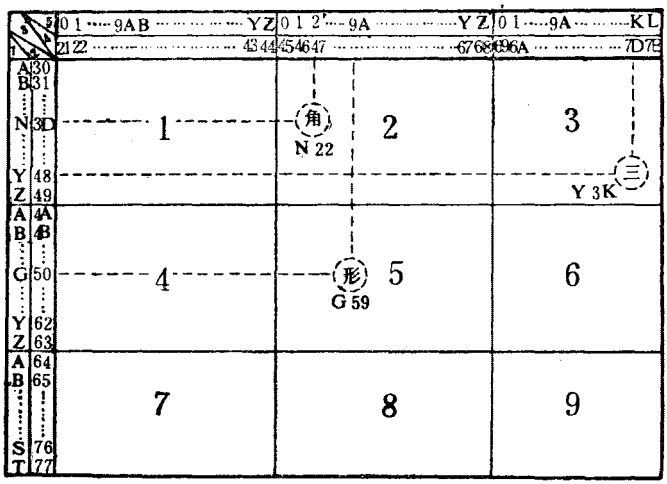


图 5·7 字母数字混合编码表

由图可见，编码表把国标码码表的汉字区（参见图5·5），重新划分成九个小区，即图5·7中的1～9区。划分的方法

是：把国标码码表纵项和横项各分成三段。纵项的三段是30～49、4A～63，64～77，分别用A～Z 26个字母来标号（见图中的①项）；横项的三段是21～44、45～68、69～7E，分别用0～9和A～Z 36个数字、字母来标号（见图中的⑤项）。

于是，国标码表中的每一个汉字可以用它所在的小区号和纵项、横项的字母数字来表示。它们的排列格式为：

纵项字母＋小区号数＋横项字母或数字

例如汉字“三”的国标码是48 7D，由于48对应纵项字母Y，7D对应横项字母K，所在小区号数是3，因此“三”字的字母数字混合码为“Y3K”。用同样方法，查出“角”字的国标码是3D47。3D对应纵项字母N，47对应横项数字2，所在小区号是2，故“角”的混合码为“N22”。再照此法查出“形”的字母数字混合码为“G59”。最后，再在混合码前加上系统自定的标识符，便构成了四字节内部码。

我们可以用以下算法将汉字的区位号直接变换成三个字节的字母数字混合码。

令 $s_1=〔(区号-30H)\div 26〕$

$s_2=\{(区号-30H)\div 26\}$

$s_3=〔(位号-21H)\div 36〕$

$s_4=\{(位号-21H)\div 36\}$

设机内码三字节为$x_1\ x_2\ x_3$

则 $x_1=s_2+41H$

$x_2=(3s_1+s_3+1)+30H$

$$x_3=\begin{cases}s_4+30H & 当s_4<10\\ s_4+37H & 当s_4\geqslant 10\end{cases}$$

其中：〔 〕为取整数值，

{ } 为取分数部分的分子值。

数值后注有H表示该数为十六进制数，

未注H的为十进制数。

有兴趣的读者，可以用上式验证一下“三角形”三字的字母数字码与查表的结果是相同的。

在得到三字节的字母数字码以后，假设仍以＆为标识符，则以四字节内部码表示“三角形ABC”的字母数字码为：

＆Y3K ＆ N22 ＆ G59 A B C,它们的代码为：

26 59334B 26 4E3232 26 473539 41 42 43

由于字母数字码的特定码型结构，一般不易混淆，故也可以用类似“方法一”的表示办法，只在汉字代码开始和结束时加标识符，即：

26 59334B 4E3232 473539 26 41 42 43这样可以缩短汉字内部码的码长，有利于提高系统处理的效率。

③ 用电报码作为汉字内部码

以电报码作为汉字传输通信码（即交换码）的部门，为了减少内部码与交换码的转换手续，常常直接把电报码作为汉字设备的内部码。

电报码是采用0～9十个阿拉伯数字给汉字编码的，因此，当汉字与阿拉伯数字混合使用时，必须使用适当的标识符将它们分隔开。通常用符号“（”表示汉字结束和阿拉伯数字开始的标识，用符号“）”表示数字结束和汉字开始的标识。

例如常见的民用电报格式：“16日13次车10车厢望接”，按照《标准电码本》译成电报码为：9916（13）2945 6508（10）6508 1666 2598 2234

其中13和10两处数字，必须用标识符（即圆括号）加以区分，“16日”在电码本中有专用代码，故不用另加括号。

计算机进行处理时，上述36个字符组成的电报码，可以用两种方法表示成机器的内部码。一种方法是把每个数字和符号都看作ASCII字符，用七位代码表编码，则其内部码字符串的数据为以下36个字节，即一个汉字4个字节：

39393136 28 3133 29 32393435 36353038 28 3130 29

36353038 31363636 32353938 32323334

另一种方法是把电报码的0～9数字作为十六进制数处理，则用两个字节便可表示一个汉字的内部码。但此时必须把电报文稿中的各种字母和符号（如上例中的括号等），都应按照《标准电码本》的编码规则译成四位数字一组的电报码。

④ 用汉字输入码作为内部码

在某些汉字系统中，为了减少从汉字输入码到汉字内部码的转换过程，直接用输入码作为汉字内部码。

例如在奥林匹亚1011中文自动打字机系统中，按照“见字识码”输入码为依据编排的汉字内部码，是以4个字母表示一个汉字，再用标识符把汉字代码与西文ASCII代码分开。

在“苹果”微机中，是以“仓颉造字法”的输入码为依据编排的汉字内部码，用1～5个小写字母表示一个汉字。由于它把七位代码表中的小写字母子集（61～7A）作为汉字代码的专用集，所以汉字代码与ASCII代码之间不必另加标识符，便能自动识别。但由于七位代码表中61～7A子集被汉字内部码占用，原西文软件中的小写字母也就不能再用了。

（2）八位码方案

八位码方案是以每一字节8个比特为基础构成的内部码。由于汉字国标码和西文ASCII码都是7位码，每个字节的最高位（b_8位）不用，始终为“0”。于是，可以用b_8位置“1”

的八位编码作为汉字内部码。这样，既能把汉字内部码与ASCII码区分开，达到中西文信息兼容；又能保持汉字内部码与国标码之间简单明确的对应关系。因此，八位码方案在微型机汉字系统和汉字终端中应用比较普遍。

汉字内部码的八位码方案大多是双字节的，即以两个八位字节来构成一个汉字内部码。双字节八位码的具体方案很多，大体上可分成两大类：双字节的最高位都置“1”的方案和双字节中第1字节的最高位置“1”的方案。

① 双字节的最高位都置“1”的汉字内部码方案

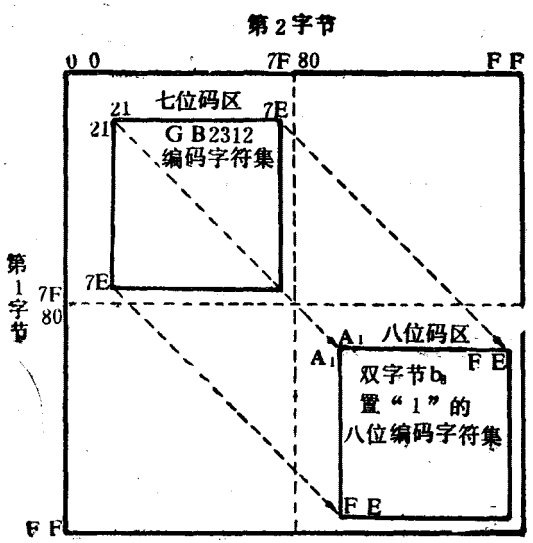


图 5·8 七位码与八位码编码字符集的映射关系

该方案是在国标《信息交换用汉字编码字符集——基本集》（GB2312—80）规定的基础上，把汉字的双字节七位码的最高位（b_8）都置为“1”，ASCII码的最高位仍置为“0”。也就是说，把汉字内部码映射到新的八位码编码字符集中，与ASCII码的七位码编码字符集完全分开。它们的映射关系如图5·8所示。

例如，汉字“统”的国标码为：

字节序号	二进制数形式	十六进制数
第一字节	×1001101	4D
第二字节	×0110011	33

其中“×”为最高位b_8，其值为“0”。

将b_8位置“1”后，汉字“统”的内部码为：

字节序号	二进制数形式	十六进制数
第一字节	11001101	CD
第二字节	10110011	B3

由图5·8可知，最高位置“1”的汉字内部码与国标码有着简明的对应关系，变换极其方便，代码长度仍为两个字节，既可以与ASCII码完全分开，又有利于提高系统处理效率。但该方案最多只能表示16384个汉字和字符图形，对某些需用汉字数量很大的系统不够用，此外对于有些把最高位为“1”的代码视为无效字符的计算机来讲，该方案就不适用了。

② 第一字节最高位恒为“1”的汉字内部码方案

把汉字代码双字节的第1字节最高位始终置“1”，第2字节的任意位都可为“1”或为“0”，与双字节最高位同时置“1”的内部码方案相比，可以使码表容量扩大一倍，能表示32768个汉字和图形字符代码。但由于第二字节的最高位并不是始终置“1”，还会出现第二字节与ASCII代码相混淆的情况，必须采取措施加以区分。

这里介绍一种第一字节最高位置“1”的内部码方案，称为移位式内部码。图5·9所示移位式内部码与国标码的码区分布关系。

Retro Workshop
果粉工作室

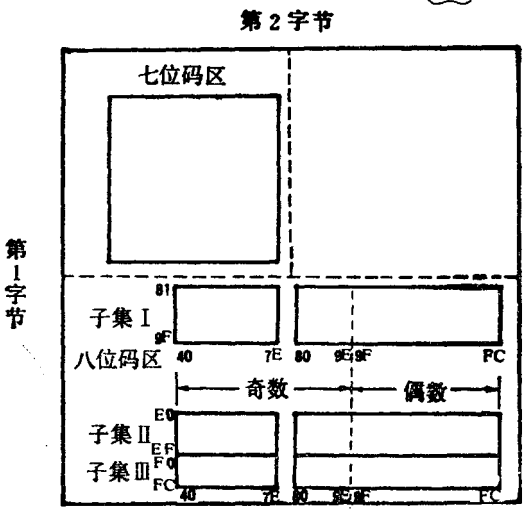


图 5·9　国标码与移位式内部码码区分布关系

图中左上角为国标码七位码区，下半部为移位式内部码八位码区。它由3个子集组成，子集Ⅰ为第一字节81～9F，共31个区；子集Ⅱ为第一字节E0～EF，共16个区；子集Ⅲ为第一

字节F0～FC，共13个区。用子集Ⅰ和Ⅱ与国标码表中的汉字区相对应，子集Ⅲ是增补字符区。移位式内部码的第二字节有188个位，分成两段。第一段40～7E共63位，第二段80～FC共125位。其中，第一段位和第二段的80～9E位为“奇数”，对应安排国标码区号为奇数的94个汉字代码；第二段的9F～FC位为“偶数”，对应安排国标码区号为偶数的94个汉字代码。这样，移位式内部码集的子集Ⅰ和Ⅱ的47个区、每区188位，正好与国标码集的94个区、每区94位一一对应。它们之间的互相转换，只需通过简单的加减和移位操作便可实现。

表5·3是“薄、病、场”三个汉字的示例：

表5·3

汉　字	国标码	区　　号	移位式内部码
薄	3121	奇数	B940
病	3221	偶数	B99F
场	3321	奇数	BA40

从上例可知，移位式内部码的第一字节最高位一定是“1”，第二字节的最高位可以是“1”或者“0”，但是它避开了ASCII码表中3F以前的各种符号及阿拉伯数字，可以提高系统对汉字代码与ASCII字符代码的鉴别能力。

但是，在进行汉字及字符信息传送时，或者在运行某些特殊的软件时，往往需要用一个字节的最高位作为校验位；或者因该位已被系统占用不允许该位为1。在这两种情况下，最高位为1的方案就不适用了。

当前，国内和国外许多汉字计算机的研究机构和制造厂商，针对它自己的计算机操作系统选用着不同的机器内部码。这些内部码各有特色和局限性。他们都迫切地期望我国能尽快

提出一套规范化、标准化的内部码方案来。我们相信，随着汉字信息处理技术的发展，我国的专家们一定会在各种机型上实现一种崭新的、功能强、效率高和兼容性好的汉字内部码。

5·2 汉字控制功能码

在汉字终端的代码体系中，有两类起着不同作用的代码。一类是用来表示汉字和非汉字图形字符的图形字符码，以上讨论的汉字内部码、交换码、字形点阵数据码、输入码等，都属于此类；另一类是用来表示汉字传输功能和设备控制功能的代码，称为汉字控制功能码。汉字控制功能码并不表示某个具体的字符图形信息，而是表示对这些字符图形的格式处理、传输控制或解释执行等过程的某种控制功能，它贯穿于汉字输入、机内处理、汉字输出和传输交换等整个处理流程。因此，汉字控制功能码的设计和选择对整个汉字系统的效率和性能有直接影响。

和图形字符码一样，汉字控制功能码也是与七位标准代码集密切相关的。七位代码集C_0集中已有的控制功能字符对于汉字设备完全适用。但必须加以扩充，以满足汉字控制功能的特殊需要。本节先介绍七位标准代码集的控制功能字符以及七位代码的扩充方法，然后讨论汉字控制功能码标准及其应用。

1. 七位标准代码表中的控制功能字符——C_0集内容介绍

“控制功能”对于许多读者并不是一个陌生的概念，在电传打字机中的“回车”、“换行”、“字母位”、“数字位”等，都属于控制功能。在通信时，发送端除了传送代表信息

内容的字符代码外，还必须把这些控制功能字符代码同时发送出去，才能使接收端的电传打字机打印出同样内容和格式的文稿来。但是电传机是用五位代码表，容纳的控制功能字符数量有限，随着信息处理技术的发展，对信息传输和设备控制的要求越来越多，于是就有了七位编码字符集的制订和应用。

七位编码字符集的代码表可以表示 $2^7=128$ 个不同的字符代码，其中94个是图形字符代码，用来表示大、小写字母、数字以及一些常用符号，也是用来扩充汉字编码字符集的基础，我们已经在上节讨论了有关这方面的内容；另外还有34个控制功能字符代码，用来表示信息传输和处理中常用的基本控制功能，这就是本节首先要讨论的内容。

这34个控制功能字符就是七位代码表中的C_0集和第 2 列第 1 个“SP”字符、第 7 列第15个“DEL”字符(参看图5·2)，大体可分成以下五类：

- 10个传输控制字符（TC_1～TC_{10}）；
- 6 个格式控制字符（FE_0～FE_5）；
- 4 个设备控制字符（DC_1～DC_4）；
- 4 个信息分隔控制字符（IS_1～IS_4）；
- 10个其它控制字符（包括SP，DEL）。

它们在七位代码表的C_0集中位组分配如表5·4所示。

下面简要介绍各类控制字符的功能和使用方法。

（1）传输控制字符TC_1～TC_{10}

传输控制字符是用来确定传输序列中汉字数据信息的性质或实现传输监控的功能字符。按其不同用途，10个传输控制字符，大体上可分为以下三种：

① 用来确定传输序列中信息性质的有 4 个，即TC_1、TC_2、

表5·4　　C_0集的控制字符分类

行＼列	0 列	1 列
0	NUL	TC_7
1	TC_1～TC_6	DC_1～DC_4
2		
3		
4		
5		TC_8～TC_{10}
6		
7	BEL	
8	FE_0～FE_5	CAN
9		EM
10		SUB
11		ESC
12		IS_1～IS_4
13		
14	SO	
15	SI	

TC_8、TC_{10}。它们的名称和作用是：

• TC_1是标题开始（SOH）字符，用作信息文电标题的第一个字符。

• TC_2是正文开始（STX）字符，用作引出正文内容，并结束标题。

• TC_3是正文结束（ETX）字符，用作结束正文内容，但

并不是结束传输。

• TC_{10}是码组结束（ETB）字符，用作结束码组，并不是结束正文。正文可以由若干个码组组成，各个码组间用ETB区分。当正文中最后一个码组结束时，应该用ETX，不是用ETB。

② 用来监控传输过程的传输控制字符有4个，即TC_5、TC_4、TC_6、TC_8。它们的名称和作用是：

• TC_5是询问（ENQ）字符，是用于请求对方给出应答的传输控制字符，这种应答可以包含站的标识（名称）和状态。在公共交换网上，建立连接后首次使用的ENQ，具有“你是谁”的含义。在其它情况下，ENQ表示要求标识和状态，并可使数据链路结束等待状态。

• TC_4是传输结束（EOT）字符，它用作表示一个或多个正文的传输结束。主站发送EOT，表示结束或放弃这次传输，从而使双方处于等待状态；从站发送EOT，表示无能力再接收信息文电。EOT也可作为对“询问”（ENQ）的否定应答。

• TC_6是确认（ACK）字符，用作对ENQ字符的肯定应答，并表示已经做好接收信息的准备。

• TC_8是否认（NAK）字符，用作接收站对发送站否定应答的传输控制字符，表示无力接收信息；或者表示接收信息有错误，并准备再接收同一电文或码组。

③ 两个特殊的传输控制字符

• TC_7是数据链转义（DLE）字符，它的作用是使紧随其后的字符或代码改变含义。用数据链转义（DLE）字符开头的字符序列，称为数据链转义序列，或DLE序列。DLE序列仅用于扩充传输控制功能。也就是说，当以上列举的传输控制

字符不够用，需增加新的传输控制功能字符时，就用DLE序列进行扩充。例如，双方传输结束，需要拆除通信线路的连接。但是上述的传输控制字符里没有“拆线”这一控制功能，于是用DLE序列扩充，规定“DLE EOT”来表示“拆线”。“DLE EOT”就是一个DLE序列，这里的“EOT”（传输结束）已经改变了它原来的含义。

• TC_0是同步空转（SYN）字符，它用于在同步传输时实现数据设备之间的同步校准。特别在设备不传输其它字符的空闲情况下，“SYN”可用来作为“时间填充”，以实现或保持字符同步。

关于DLE以及其它传输控制字符的具体用法，将在5·3节有关汉字通信的链路传输控制规程中介绍。

（2）格式控制字符FE_0～FE_3

格式控制字符是在文字图形打印或显示输出时，用于改变其记录位置的控制功能字符。七位代码表C_0集中有6个格式控制字符，它们的名称、缩写符及控制功能是：

• FE_0是退格（BS），用于文字图形的打印或显示位置在同一行内反向移动一个格位。

• FE_1是横向制表（HT），使打印或显示位置进至同一行内下一组预定的格位。

• FE_2是换行（LF），使打印或显示位置进至下一行同一格位。若经过使用双方约定，也可进至下一行的第一个格位，即具有新行（NL）的含义。

• FE_3是纵向制表（VT），使打印或显示位置进到下一组预定行同一格位。若经过双方约定，也可以进至下一组预定行的第一个格位。

• FE_4是换页（FF），使打印或显示位置进至下一表格或下一页的预定行上同一格位。若经过双方约定，也可进至预定行上的第一个格位。

• FE_5是回车（CR），用于使打印或显示位置退回至同一行第一个格位。

以上是六个基本的格式控制字符，不论在西文设备或汉字设备中都经常使用。在汉字文本打印输出时需用的其它格式控制功能，将在有关标准中增补和扩充。

（3）设备控制字符DC_1～DC_4

设备控制字符是用于控制同数据处理系统或通信系统连接的辅助设备的控制字符。在七位代码表中，设备控制字符有4个（DC_1、DC_2、DC_3、DC_4），主要用于辅助设备的启动和停止、设备的基本操作方式和特定操作方式之间的转换控制。

• DC_1、DC_2主要用于使辅助设备接通或者启动。在有两台辅助设备的情况下，可以用DC_1使一台设备恢复其基本操作方式；用DC_2使另一台设备建立其特定的操作方式，这时，仍用DC_1来使它恢复到基本操作方式。

• DC_3、DC_4主要用于使辅助设备断开、停止或中断。DC_3是使设备处于"准关闭"状态，例如等待、暂停、挂起等。这时，也是用DC_1来恢复它的正常操作。

设备控制字符与传输控制字符的区别在于：前者只用于控制同通信系统连接的辅助设备，不能控制系统本身。控制通信系统的启动（接通）或停止（断开）应使用传输控制字符。

设备控制字符的通常用法是：使用一台辅助设备时，用DC_2启动，DC_4停止；使用两台辅助设备时（如一台辅助接收器、一台辅助发送器）用DC_2和DC_4来控制辅助接收设备的启

动和停止；用DC_1和DC_3来控制辅助发送设备的启动和停止。

（4）信息分隔字符IS_1～IS_4

信息分隔字符是用于在数据的分层次排列中，从逻辑上将数据加以分隔或限定的控制字符。在七位代码表中，有4个信息分隔字符，当分级使用时，将它们从低阶到高阶的排列顺序是：IS_1、IS_2、IS_3、IS_4。在分级使用中被某个分隔字符所限定的数据，不能再用高阶分隔符将它分开，而只能被高阶分隔符限定。现将信息分隔符的名称、缩写符和作用分述如下：

·IS_1是单元分隔字符（US），被它所限定的数据项称为一个“单元”。

·IS_2是记录分隔字符（RS），被它所限定的数据项称为一个“记录”。

·IS_3是群分隔字符（GS），被它所限定的数据项称为一个“群”。

·IS_4是文卷分隔字符（FS），被它所限定的数据项称为一个“文卷”。

当信息分隔符不分级使用时，对它们的各种讲法，必须分别规定其特定的含义。

（5）其它控制字符

其它控制字符包括了在七位代码表G_0集中除上述四类控制字符以外的8个特殊控制字符，还包括2个具有控制功能的间隔（SP）和抹掉（DEL）字符。它们的名称、缩写符及含义如下：

① 空白字符（NUL） 是一个以全“0”为代码的特殊控制字符，用于实现媒体填空或时间占空。从汉字或字符

串中插入或去掉空白字符，均不影响该汉字或字符串的含义，而只对设备或格式起到控制作用。

② 告警字符（BEL） 当需要提醒设备操作人员注意时所使用的一个控制字符，用它可以控制警报器，或发出一个清晰的铃声。

③ 移出字符（SO）

④ 移入字符（SI）

这是一对用来扩大图形字符集的特殊控制字符。移出字符（SO）使紧随其后、直到出现移入字符（SI）为止的一串字符中，凡属七位代码表G_0集的图形字符都可能被改变含义，仅间隔（SP）和抹掉（DEL）不受影响。移入字符（SI）结束了由移出字符引进的字符串、并使其后的字符恢复它在七位代码表中的含义。现以图5·10来说明SO和SI的使用方法。如图示，在某汉字终端中的字符串被SO和SI字符分隔成三段，

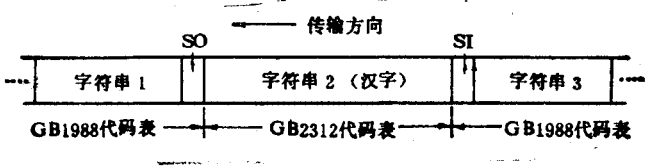


图 5·10 SO和SI控制字符的用法

即字符串1、字符串2和字符串3。已知该终端使用两种代码表：七位标准代码表和GB2312代码表，则按照字符串传送方向，串1和串3中的图形字符代码属于七位标准代码表的范围，串2中的图形字符代码属于汉字国标码GB 2312代码表的范围。

⑤ 转义字符（ESC） 这是用来扩充控制功能（不包

括传输控制功能TC_1～TC_{10}）的特殊控制字符。它使紧随其后的有限个字符代码改变原有的含义，而是以约定的另外一些字符代码解释。

ESC字符与DLE字符的用法是不同的。DLE字符仅用于扩充数据传输控制功能，它的用法在有关传输控制规程（如GB3453）中规定（见5·3节）。ESC字符用于扩充七位编码字符集，它的具体用法由《信息处理交换用七位编码字符集的扩充方法》（GB2311—80）规定，我们将在本节的下一小节中介绍。

⑥ 作废字符（CAN） 用于按照某个约定指明哪些与它有关的数据是错误的，或是可忽略的。使用作废字符时，必须规定它的特定含义。

⑦ 取代字符（SUB） 用于替换被认为是无效或错误的字符，或者替换在给定的设备上不能表示的字符。

⑧ 媒体结束字符（EM） 用于识别数据媒体的物理末端或媒体上数据占用部分的末端，如凿孔纸带或记录磁带的尽头等。

⑨ 间隔字符（SP） 用来使打印或显示位置在同一行内前进一个格位（字距），通常是以在一串字符中空出一个格拉来表示。间隔字符不在七位代码表的C_0集内，但它同时具有格式控制和信息分隔控制字符的功能。

⑩ 抹掉字符（DEL） 也不在C_0集内，是一个以全“1”为代码的控制字符，主要用于抹掉凿孔纸带上错误的或不需要的字符（使纸带一排全部有孔）。与NUL字符一样，DEL字符也可以用于实现媒体填空或时间占空。从汉字或字符串中去掉或插入DEL字符，均不影响汉字或字符串的含义，只对设备或格式起控制作用。

现将34个控制字符的名称及功能分类列表见表5·5。

表5·5　　控制功能字符分类表

分类	名　称	缩　写	代　码	功　能
传输控制字符TC	标题开始	SOH	0/1	用于确定传输序列中信息的性质
	正文开始	STX	0/2	
	正文结束	ETX	0/3	
	码组结束	ETB	1/7	
	询　问	ENQ	0/5	用于正向或反向监控序列
	确　认	ACK	0/6	
	否　认	NAK	1/5	
	传输结束	EOT	0/4	
	同步空转	SYN	1/6	用于设备间的同步或时间填空
	数据链转义	DLE	1/0	用于扩充传输控制功能
格式控制字符FE	退　格	BS	0/8	使打印在同一行内移动一个格位
	横向制表	HT	0/9	使打印进至本行内下一组预定的格位
	换　行	LF	0/10	使打印进至下一行的同一格位
	纵向制表	VT	0/11	使打印进至下一组预定行的同一格位
	换　页	FF	0/12	使打印进至下一页预定行的同一格位
	回　车	CR	0/13	使打印退至本行第一个格位
设备控制字符DC	设备控制1	DC1	1/1	使辅助设备启动或接通
	设备控制2	DC2	1/2	
	设备控制3	DC3	1/3	使辅助设备停止或断开
	设备控制4	DC4	1/4	

续表

分类	名称	缩写	代码	功能
信息分隔字符IS	单元分隔	US	1/15	限定或分隔数据项
	记录分隔	RS	1/14	
	群分隔	GS	1/13	
	文卷分隔	FS	1/12	
其它控制字符	空白	NUL	0/0	用于实现媒体填空或时间占空
	告警	BEL	0/7	提醒人们注意
	移出	SO	0/14	改变随后的字符串的原有含义
	移入	SI	0/15	结束移出字符功能，恢复原有含义
	转义	ESC	1/11	扩充控制功能
	作废	CAN	1/8	指出错误的或可忽略的数据
	取代	SUB	1/10	替换错误的或无效的字符
	媒体结束	EM	1/9	识别媒体的物理末端
	间隔	SP	2/0	使打印在同一行中前进一个格位
	抹掉	DEL	7/15	使凿孔纸带全部有孔以抹掉错误

2. 控制功能代码的扩充
——国标GB2311-80内容简介

在实际应用中，GB1988-80《信息处理交换用的七位编码字符集》（即七位代码表）规定的128个字符往往不能满足使用需要，于是要进行代码扩充。国标GB2311-80《信息处理交换用七位编码字符集的扩充方法》是参照国际标准ISO DIS 2022制订的，它用于七位代码系内，扩充其控制功能或

（和）图形字符；也可用到八位代码系中。本小节简要介绍GB2311-80中有关控制功能代码的扩充方法。

（1）控制功能码的扩充元素

控制功能码的扩充元素是指允许扩充的控制功能字符或控制字符集。它们是系统设计者根据需要预先制定的字符集，可以用规定的方法去指明和调用。

控制功能码的扩充元素有三种类型：C_0集、C_1集和单个增补的控制功能。

① C_0集

C_0集是由32个控制字符构成的控制字符集，它应包含ESC、SI、SO三个控制字符，并保持其在GB1988-80七位代码表中的位置和意义不变。C_0集中还可包含10个传输控制字符和NUL（空白）、SUB（取代）、CAN（作废）三个控制字符，它们在七位代码表中的位置和意义也应保持不变。C_0集内的控制字符代码都用单个位组（即单字节）表示，但需用三字节的“ESC 2/1 F”序列来指明和调用。

② C_1集

C_1集是由32个控制功能构成的增补控制功能集，它的控制功能代码都是用2个位组（即双字节的ESC Fe序列）来表示，用ESC 2/2 F来指明和调用。

C_1集内不应包括传输控制功能。

③ 单个增补的控制功能

单个增补的控制功能是指不包含在C_0集和C_1集内的增补控制功能，用ESC 2/3 F序列来表示，其功能由终止字符F决定。

（2）转义序列的构成和作用

上面提到的ESC 2/1 F、ESC Fe等以ESC开头的字符序列，称为转义序列。它的作用是在进行代码扩充时去指明、调用或表示某一特定的扩充元素，实现代码扩充。

转义序列的一般格式是：

ESC I F

其中，I称为中间字符，可以取七位代码表中第2列的一个或几个位组；F称为终止字符，可以取七位代码表中第3列到第7列中除7/15以外的任意一个位组。它们的区域分配如图5·11所示。中间字符和终止字符在分配区域内的不同取值，表示转义序列的不同含义。在转义序列中，中间字符可以省略，

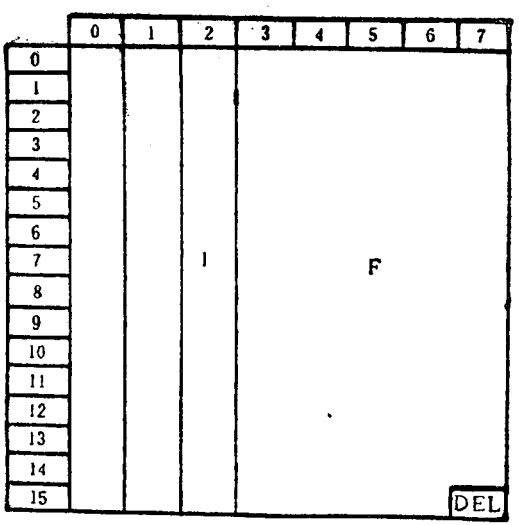


图 5·11 转义序列的中间字符和终止字符的区域分配

终止字符还用来表示转义序列的结束，不能省略。

用转义序列进行代码扩充的方法，是使跟随在ESC字符之后的中间字符和终止字符有不同的取值。这时，这两个字符所表示的意义与它们单独出现时的意义无关。

用于扩充控制功能的转义序列，主要有二字符转义序列和三字符转义序列两种：

① 二字符转义序列

二字符转义序列是指省略了中间字符 I 的转义序列，它的一般格式是：

ESC F

二字符转义序列共有79个，用以表示增补的控制功能。根据终止字符F在七位代码表中的不同位置，可分成三类（见图5·12）：

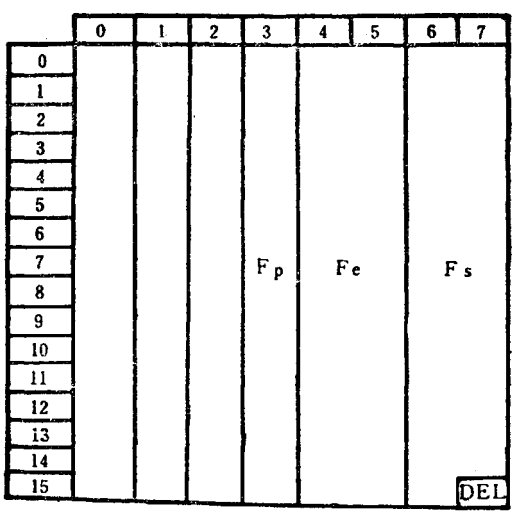


图 5·12 二字符转义序列的终止字符分配

如图所示，占用第3列的终止字符称为Fp，占用第4、5列的终止字符称为Fe，占用第6、7列的终止字符称为Fs。由它们构成的转义序列分别表示的不同含义是：

·ESC Fp表示专用的单个增补的控制功能，其功能由信息交换双方预先约定。

·ESC Fe表示标准的、由32个控制功能组成的C_1集，其功能由终止字符决定。

·ESC Fs表示标准的单个增补的控制功能，其功能由终止字符决定。

如前所述，对于扩充的C_0集的控制功能，仍由单个位组的功能字符表示，但需用三字符的转义序列ESC 2/1 F指明和调用。

② 三字符转义序列

三字符转义序列是由一个转义字符ESC、一个中间字符I和一个终止字符F组成的转义序列，它的序列形式是ESC I F。其中I和F在七位代码表中的位置如图5·13所示。

三字符转义序列的含义由中间字符I决定，当用于扩充控制功能时，I取位组2/1、2/2、2/3，它们的具体含义是：

·ESC 2/3 F用于表示单个增补的控制功能，其功能由终止字符决定。

·ESC 2/1 F用于指明和调用C_0集。该集由七位代码表0列和1列的位组所表示的32个控制字符组成。

·ESC 2/2 F用于指明和调用C_1集，但不影响C_0集的使用。如前所述，C_1集是由32个控制功能构成的增补控制功能集，它的每一个控制功能代码都是用二字符转义序列ESC Fe表示。

三字符转义序列终止字符F分为Fp和Ft两种（见图5·13）。Ft占用七位代码表的4～7列（7/15除外），共有63个，作为

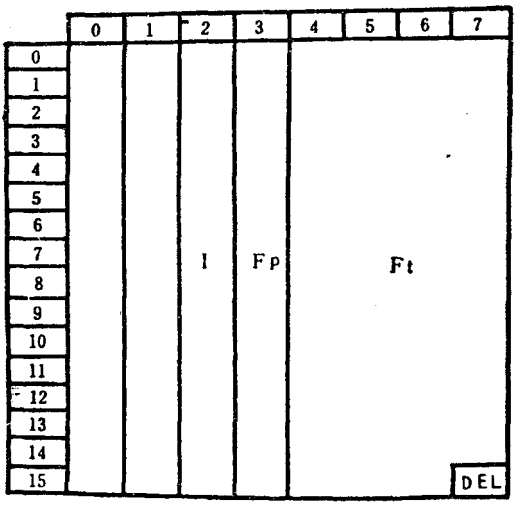


图 5·13 三字符转义序列的中间字符和终止字符的分配

标准的代码扩充使用；Fp占用七位代码表第 3 列，共有16 个，留作专用，允许系统设计者自由使用，不具有标准意义。

这里说明一下“指明”、“调用”和“表示”的不同含义。“指明”是在预先制订的若干个字符集中用规定的方法指出其中某一个字符集。“调用”是使已指明的字符集成为能使用的状态。“表示”是在已指明和调用的字符集中，按照字符或控制功能的意义，使用规定的位组或转义序列。它们的具体应用，将在下文中举例说明。

（3）控制功能的扩充方法

在实际工作中，如遇七位代码表中的控制功能不够用，需要扩充时，可以采用以下两种方法扩充控制功能。

① 用转义序列扩充控制功能

用转义序列扩充控制功能的方法是：先将需要用的控制功能构成若干个C_0集、C_1集或单个增补的控制功能，使用时以下列转义序列分别指明、调用和表示：

·以ESC 2/1 F指明和调用若干个C_0集内的一个C_0集，再用0列或1列中的一个位组表示该C_0集内的一个控制功能字符。

·以ESC 2/2 F指明和调用若干个C_1集内的一个C_1集，再用4列和5列中的位组构成的二字符转义序列ESC Fe来表示该C_1集内的控制功能。

·用ESC Fs、ESC Fp、ESC 2/3 F来表示单个增补的控制功能，其功能由终止字符确定。

图5·14是用转义序列扩充控制功能代码的扩充方法示意图。

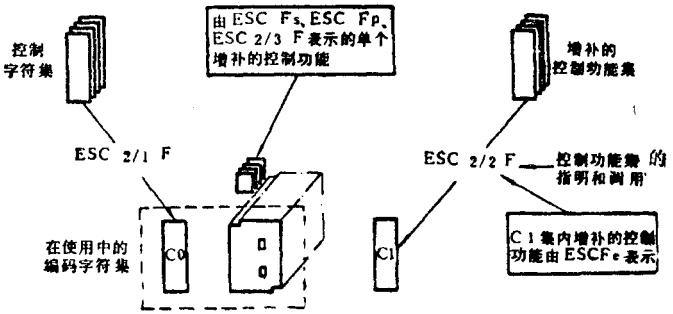


图 5·14 用转义序列扩充控制功能的扩充方法

采用转义序列扩充控制功能的方法，规则比较明确一致，控制功能的扩充范围大，可以满足任何系统对控制功能的扩充需要。例如仅用三字符序列ESC 2/2 F指明和调用C_1集的方

法，便可扩充增补的控制功能一千多个（可有63个C_1集，每集可包括32个控制功能）。但是，用转义序列的扩充方法每次都要用一个三字符转义序列来指明和调用，给使用带来很大不便。尤其对绝大多数汉字终端和汉字微机系统来讲，不需扩充这么多控制功能，因此，人们又设计了一些比较简便实用的控制功能扩充方法，如用控制序列引导符的扩充方法，就是常用的一种。

② 用控制序列引导符扩充控制功能

如上所述，为了适应大多数设备扩充控制功能的需要，国际标准化组织(ISO)已经在1983年为字符成形设备选用了87个增补的控制功能和18种功能配置模式。并已制定了相应的国际标准《字符图形设备的附加控制功能》(ISO/6429—1983)。该标准规定的控制功能扩充方法是：设一个增补的控制功能集C_1集，在C_1集内设有一个特殊的“控制序列引导符”CSI，再由CSI序列来指明和调用其它增补的控制功能（见图5·15）。这就是用控制序列引导符的控制功能扩充方法。

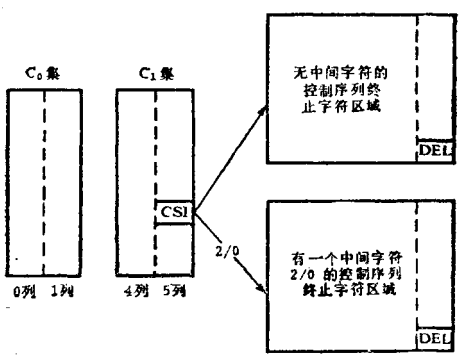


图 5·15 用CSI序列扩充控制字符的扩充方法

用控制序列（即CSI序列）扩充控制功能的通用格式为：

CSI $P_1 \cdots P_n$ $I_1 \cdots I_m$ F

其中，CSI为控制序列引导符，由位组5/11表示；$P_1 \cdots P_n$为参数值，由3列中的位组表示；$I_1 \cdots I_m$为中间字符，由2列中的位组表示；F为终止字符，由4～7列中的位组表示(7/15除外)。

用控制序列引导符的扩充方法，是在符合GB2311-80关于转义序列扩充方法和扩充规则的前提下，避免了由于使用多个C_1集和C_0集而带来的转义序列过长的缺点，使用起来比较方便，但用此法扩充的控制功能数量有限。为了便于国际间进行统一的信息交换，我国1985年制订的《文字和符号成形设备的增补控制功能》(GB5261—85)，也采用了这种扩充方法。

3. 文字和字符成形设备的增补控制功能——国标GB5261—85内容简介

在计算机信息处理系统和通信系统中，对终端设备的要求，除了能正确表达信息的内容外，还要能表达信息的形式。随着终端控制功能的不断增加，七位代码表C_0集内的34个控制字符已不能满足要求，尤其是汉字终端设备的显示、打印文本格式处理方面，与西文终端设备相比还有不少特殊的使用要求。例如：在文字行的排列方式上，西文设备基本上是横排格式输出，汉字设备则需要横排和直排两种格式输出；又如“外字”处理方面和中西文混合格式处理方面，西文设备基本上不存在什么问题，汉字设备则不同，虽然用了容量达数兆位的字库，仍然难免会碰到没有收入的字形需要进行“外字”处理，如造字、点阵传送和装配等。我国制订的国家标准《文字和符号成形设备用的增补控制功能》GB5261—85，是在国际标准ISO/6429—1983的基础上，进行了必要的修改和补充，共增

加了处理汉字信息的14种控制功能和3种模式。使这个国家标准既适用于汉字信息处理和文本输出，又保持与国际标准充分兼容。

所谓“字符成形设备”是指能发送和接收由控制功能和图形字符组成的编码数据流，并能产生字符图形输出，以供人们阅读的设备。

GB5261—85依据GB2311—80《信息处理交换用七位编码字符集的扩充方法》的规定，增补了包括一个C_1集及其引伸出来的控制功能以及若干单个控制功能。当这些控制功能夹在进行交换的字符编码数据中时，它们可以与GB1988—80《信息处理交换用的七位编码字符集》中的C_0集配合使用。

现将GB5261—85《文字和符号成形设备的增补控制功能》的名称分类、编码表示方法，以及若干常用控制功能的应用举例简要介绍如下。

（1）控制功能的名称和分类

在GB5261中，增补控制功能共有101种，按用途可分为以下10类：

① 控制串定义符6个

- APC 应用程序命令
- OSC 操作系统命令
- CPT(s…) 字形点阵传送
- PM 保密消息
- DCS 设备控制串
- ST 串终止符

② 引导符2个

- CSI 控制序列引导符
- ICSI 汉字控制序列引导符

③ 转移功能2个

- SS2 单移2
- SS3 单移3

④ 格式控制符22个

- HPA(n) 字向位置绝对移动
- HPB(n) 字向位置反向移动

- HPR(n) 字向位置正向移动
- HTJ 对齐的字向制表
- HTS 置字向制表
- HTSA(n…) 置字向制表绝对位置
- HVP(n,m) 置字向和行向位置
- IND 移行
- INL(n) 汉字新行
- NEL 下一行
- PLD 行部分下移
- PLU 行部分上移
- PPA(n) 页位置绝对移动
- PPB(n) 页位置反向移动
- PPR(n) 页位置正向移动
- RI 反向移行
- SGR(s…) 图形显现式样选择
- TBC(s…) 清制表
- VPA(n) 行向位置绝对移动
- VPB(n) 行向位置反向移动
- VPR(n) 行向位置正向移动
- VPS 置行向制表

⑤ 文本排版设备用的增补格式控制符14个

- CRP 字符旋转处理
- EPN 注音结束
- FNT(s;t) 字体选择
- GSM(n;m) 字符图形大小修改
- GSS(n) 字符图形大小选择
- HSS(n) 半字指定
- JFY(s) 整版
- LLS(n) 行长指定
- QUAD(s…) 单行对齐
- SPD(s…) 书写方向选择
- SPI(n;m) 间隔指定
- SPN 注音开始
- SSU(s…) 单位大小选择
- TSS (n) 窄间隔指定

⑥ 修改可见数据用的编辑功能9个

- DCH(n) 删字
- DL(n) 删行
- EA(s…) 消区
- ECH(n) 消字
- ED(s…) 消页
- EF(s…) 消段
- EL(s…) 消行
- ICH(n) 插字
- IL(n) 插行

⑦ 移动操作位置用的控制功能18个

- CBT(n) 位标反向制表
- CHA(n) 位标字向绝对移动
- CHT(n) 位标字向制表
- CNL(n) 位标正向移行
- CPL(n) 位标反向移行
- CTC(s…) 位标制表控制
- CUB(n) 位标反向移动
- CUD(n) 位标行向正移
- CUF(n) 位标正向移动
- CUP(n;m) 置位标
- CUU(n) 位标行向反移
- CVT(n) 位标行向制表
- NP(n) 正向换页
- PP(n) 反向换页

- SD(n) 下滚
- SL(n) 左滚
- SR(n) 右滚
- SU(n) 上滚

⑧ 区的限定5个

- DAQ(s…) 限定区定义符
- EPA 防保区结束
- ESA 选定区结束
- SPA 防保区开始
- SSA 选定区开始

⑨ 模式设置2个

- RM(s…) 清模式
- SM(s…) 置模式

⑩ 其他控制功能21个

- CCH 字符作废
- CPR(n;m) 位标位置报告
- DA(s…) 设备属性
- DMI(Fs) 禁止人工输入
- DSR(s…) 设备状态报告
- ECC 字符合成结束
- EMI(Fs) 允许人工输入
- ECCI(s…) 外字代码指示
- IDCS(s…) 设备控制串标识
- INT(Fs) 中断
- MC(s…) 媒体拷贝
- MW 消息等待
- PU1 专用1
- PU2 专用2
- REP(n) 重复
- RIS(Fs) 恢复初始状态
- SCC 字符合成开始
- SCPC(s…) 字形点阵构成选择
- SEC(s,t) 外字代码构成选择
- SEE(s…) 编辑范围选择
- STS 置发送状态

在缩写词后括号里的字母的含义：

n和m表示数值参数；s和t表示选择参数；

n…和s… 表示参数个数可变；Fs表示是ESC Fs序列；无标记的表示是C_1集内的一个控制功能，或是由ICSI序列构成的无参数控制功能。

（2）控制功能的编码表示

以上列举的101个控制功能，按照它们的不同性质，分别用以下三种编码方法表示：

① 将其中26个控制功能编成C_1集，用二字符转义序列ESC Fe表示。Fe是七位代码表中4列和5列的位组。这26个控制功能的位组分配见表5·6。

表5·6　　C_1集中控制功能的位组分配

行　号	列　号	
	4	5
0	—	DCS
1	—	PU1
2	—	PU2
3	ICSI	STS
4	IND	CCH
5	NEL	MW
6	SSA	SPA
7	ESA	EPA
8	HTS	—
9	HTJ	—
10	VTS	—
11	PLD	CSI
12	PLU	ST
13	RI	OSC
14	SS2	PM
15	SS3	APC

由表可知，控制序列引导符CSI和汉字控制序列引导符ICSI，它们的终止字符Fe的位组分别是5/11和4/3。因此，可以写出它们的编码表示是：

CSI⟶ESC　5/11

ICSI⟶ESC　4/3

表中还空着6个位组，留待今后标准化之用。

②　将其中4个控制功能编为标准的单个增补控制功能，由GB2311—80规定的ESC Fs序列表示，(Fs取七位代码表的第六、七列位组)。它们的名称、缩写形式和编码表示如下所示：

缩写形式	名　称	编　码
DMI	禁止人工输入	ESC 6/0
EMI	允许人工输入	ESC 6/2
INT	中　断	ESC 6/1
RIS	恢复初始状态	ESC 6/3

③ 其余的71个控制功能，用CSI或ICSI引导的控制序列表示，其通用的格式为：

CSI（或ICSI）$P_1 \cdots P_n$ $I_1 \cdots I_m$ F

式中符号的含义已在本节第2小节（3）、②项中说明（见第321页）。在GB5261—85中，这71个控制功能分为以下三种编码表示方法：

表5·7　无中间字符的CSI控制序列的终止字符的位组

行　号	列　号		
	4	5	6
0	ICH	DCH	HPA
1	CUU	SEE	HPR
2	CUD	CPR	REP
3	CUF	SU	DA
4	CUB	SD	VPA
5	CNL	NP	VPR
6	CPL	PP	HVP
7	CHA	CTC	TBC
8	CUP	ECH	SM
9	CHT	CVT	MC
10	ED	CBT	HPB
11	EL	—	VPB
12	IL	—	RM
13	DL	—	SGR
14	EF	—	DSR
15	EA	**SPD	DAQ

（A）用无中间字符的CSI控制序列表示的控制功能符44个。其编码格式为：

CSI F

终止字符F在七位代码表中的位组如表5·7所示。

（B）用有一个中间字符2/0的CSI控制序列表示的控制功能符15个，其编码格式为：

CSI 2/0 F

终止字符F在七位代码表中的位组如表5·8所示。

表5·8 有一个中间字符2/0的CSI控制序列的终止字符的位组

行号	列号		
	4	5	6
0	SL	PPA	
1	SR	PPR	
2	GSM	PPB	
3	GSS		
4	FNT		
5	TSS		
6	JFY		
7	SPI		
8	QUAD		
9	SSU		
10	—		
11	—		
12	—		
13	—		
14	HTSA		
15	IDCS		

（C）用ICSI控制序列表示的控制功能12个，都没有中间字符，其编码格式为：

ICSI F

终止字符F在七位代码表中的位组如表5·9所示。

表5·9 ICSI控制序列的终止字符的位组

行号	列号		
	4	5	6
0	ECC	HSS	SEC
1	SCC	LLS	
2	SPN		SCPC
3	ECCI		CPT
4	EPN		CRP
5			
6			
7			
8			
9			
10		INL	
11			
12			
13			

（3）汉字控制功能应用举例

GB5261-85的增补控制功能的内容比较多，现列举在汉字终端设备中较常用的几种控制功能，如字符合成、文稿格式处理和字形点阵传送和装配等，来说明控制功能字符序列的使用方法。

① 字符（汉字）合成

当需要打印或显示的图形字符没有存放在字形发生器中时，有时可以用“字符合成”的办法解决。

如果合成的图形字符比较简单，可以使用七位代码表中规

定的“退格（BS）”控制字符（位组0/8）实现。例如：字符发生器中没有“≠”（不等于）符号，需要输出时，可以用“=”（等号，位组3/13）和“＼”（斜线，位组5/12）加上“退格”字符来合成。即输入字符序列：“= BS＼”，设备上就可输出合成符号“≠”。

如果需合成较复杂的图形字符或汉字时，就得使用扩充的控制功能符“SCC”（字符合成开始）和“ECC”（字符合成结束）。我们从表5·9查知，SCC和ECC是用ICSI引导的无中间字符控制序列表示，终止字符的位组是4/1和4/0。也就是说，SCC的编码表示是ICSI 4/1，ECC的编码表示是ICSI 4/0。于是，当输入字符序列：“SCC田又又又又ECC”时，就是指示设备输出合成字符“畾”，字符合成的具体方法，取决于设备的执行过程。

对一些无法合成的复杂汉字和图形的输出，则需用专门的造字功能来解决。

② 文稿格式处理

输出文稿的格式是否符合使用要求，是衡量打印终端设备性能的一项重要指标。在ISO/6429的增补控制功能中，大多是有关格式处理的控制功能。汉字文稿的格式处理又有其特点。在GB5261中，为了汉字文稿的输出方便，又增补了汉字新行（INL）、字符旋转处理（CRP）、注音开始和结束（SPN、EPN）、行长指定（LLS）等十多个专为汉字格式处理用的控制功能字符。现举INL、LLS等几个功能字符的应用为例，说明汉字文稿一般格式的处理方法。

汉字新行（INL）是为汉字文稿输出时需要分段或另起一行时设置的一个控制字符。从表5·9中查知，它是由ICSI引导、带有一个数值参数n、终止字符位组为5/11的字符序列，

编码表示是：ISCI n 5/11，执行结果是使打印的操作位置移到下一行第n个字的位置。根据标准规定，n有一个缺省值（即隐含值）等于3，因为一般文稿在段落结束另起一行时，通常从下一行第3个字的位置开始。因此，如果出现编码序列ICSI 5/11，即数值参数被省掉，表示操作位置是从下一行第3个字位开始。

行长指定（LLS）是文本排版设备用的增补格式控制符。从表5·9查知，也是由ICSI引导、带一个数值参数n的控制序列，终止字符位组为5/1，编码表示是ICSI n 5/1，标准规定n没有缺省值，因此不能省略。

LLS的功能是指定文稿打印或印刷时一行的长度，其作用在出现下一个LLS字符之前一直有效。长度单位则由“单位大小选择”（SSU）确定。

SSU是确定一个控制功能参数值单位大小的格式控制字符，它是有一个中间字符2/0和一个选择参数n的CSI控制序列。从表5·8查知，终止字符位组为4/9，它的编码表示为：CSI n 2/0 4/9。标准规定参数值n没有缺省值，n值与所代表单位的对应关系为：

n值	单位
1	“国际印刷标准”
2	“0.1点”（即0.0353mm）
3	“0.1第道点”（即0.0376mm）
4	“密耳”（即0.0254mm）
5	留作以后标准化用
6	微米（即0.001mm）
7	象素（在设备中能指定的最小增量）

现以图5·16来说明格式控制字符INL、SSU和LLS的应

用。

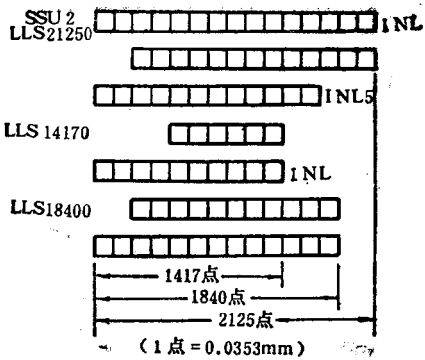


图 5·16 格式字符INL、SSU、LLS应用举例

如图所示，在文字字符串开始，用SSU 2、LLS 21250两个控制符，指明以下文字的行长为2125点（约750mm）。在打印完规定行长后，根据INL控制符，下一行从第3个字的位置开始（缺省值为3）。在没有新的LLS字符之前，仍印满原来规定行长即2125点。以下两个LLS 14170和LLS 18400控制符，使后面四行的行长分别改变成1417点和1840点，因为没有新的SSU控制符出现，行长单位一直保持为“0.1点”。在第三行末出现的INL 5，指示设备在下一行从第5个字的位置上开始打印；第五行末的INL指示设备在下一行第3个字位打印。没有出现INL时，设备在按规定行长打满一行后，从下一行第一个字的位置开始打印。第三行没有打印到规定行长，是表示该段内容已经打印完了。

③ 字形点阵传送和装配

现用的汉字终端设备绝大多数以GB2312-80汉字基本集内的汉字及字符图形作为汉字库容量的依据。遇到需要基本集以外的字形打印显示输出而又无法用字符合成功能时，则用系统

的“造字功能”来解决。为了使本机所造的字形能在远端的设备上正确地显现出来，就需要应用“字形点阵传送和装配”功能，这是汉字设备之间进行信息交换时经常遇到的一个问题。

为了实现上述功能，GB5261增补了“字形点阵传送”（CPT）和“字形点阵构成方式”（SCP）两种控制功能符。

字形点阵传送（CPT）是由汉字控制序列引导符ICSI引导的、无中间字符的控制序列来表示，终止字符的位组为6/3（见表5·9）。其编码表示为：ICSI 10 6/3。其中10是它的选择参数，表明所传送的字形点阵是汉字编码集以外的“外字”。

CPT的功能是指明其后续的数据串是一组汉字或图形字符代码及其字形点阵数据，并以串终止符（ST）结束。例如，有字符数据串：

CPT $a_0a_1a_2a_3a_4a_5\cdots\cdots a_{86}a_{87}$ ST

则a_0～a_3是指汉字或图形字符代码，a_4～a_{87}是字形点阵数据，并用GB1988-80七位代码表中第3列位组的低4位（即$b_4b_3b_2b_1$）表示。

当传送的字符数据串为以下格式：

CPT $a_0a_1a_2a_3$ ST

则表示向对方索要代码为$a_0a_1a_2a_3$的字形点阵。

字形点阵构成选择（SCPC）也是以ICSI引导的无中间字符的控制序列表示，终止字符的位组为6/2（见表5·9）。它带有四个选择参数，其通用格式为：

ICSI s_1, s_2, s_3, s_4 6/2

其中，s_1——指定横向点数1…m点，缺省值为1

s_2——指定竖向点数1…n点，缺省值为1

s_3——指定扫描方法：0表示横向扫描（缺省值）

1 表示竖向扫描

s_4——指定装配方法：0 表示非连续装配（缺省值）

1 表示连续装配

SCPC的功能是指明字形点阵的构成（装配）方法。在发送方传送字形点阵数据以后，应随即传送SCPC控制字符，否则接收方就无法装配出正确的字形来。

在本例中，当发送上述第一个CPT序列以后，应即发送ICSI 15 16 6/2序列。该ICSI序列指明在CPT序列中的a_4～a_{67}64个数据是用于构成15（横向）×16(竖向)字形点阵的32个字节。后两个参数s_3和s_4缺省，表示是用横向扫描、非连续装配的方式。关于横向扫描和竖向扫描方式问题，我们已经在第四章中讨论过。这里讨论一下关于连续装配和非连续装配问题。

连续装配与非连续装配是接收方在收到CPT序列字符数据串以后，把其中a_4以后的数据装配成点阵字形的两种不同方法，由CPT序列中的参数s_4指明。当s_4为“1”时表示是连续装配；s_4为“0”或者缺省时，表示非连续装配。

所谓连续装配就是把收到的数据以从左到右、从上到下的扫描方向连续不断地装配成字形点阵。我们以图5·17所示的15×16横排格式字形点阵为例来说明。

图中的小圆点是构成字形的许多点子，对于15×16点阵字形共有240个点位，小圆中间的标号是以连续扫描顺序编的号数。当采用连续装配方式时，每4个点位对应一个16进制数a_n，即1～4点位对应a_4，5～8点位对应a_5，……237～240点位对应a_{63}。也就是说，用a_4～a_{63}共60个16进制数（即30字节）就可以装配出一个15×16点阵字形。

所谓非连续装配就是当每一个新扫描行开始时，都用一个

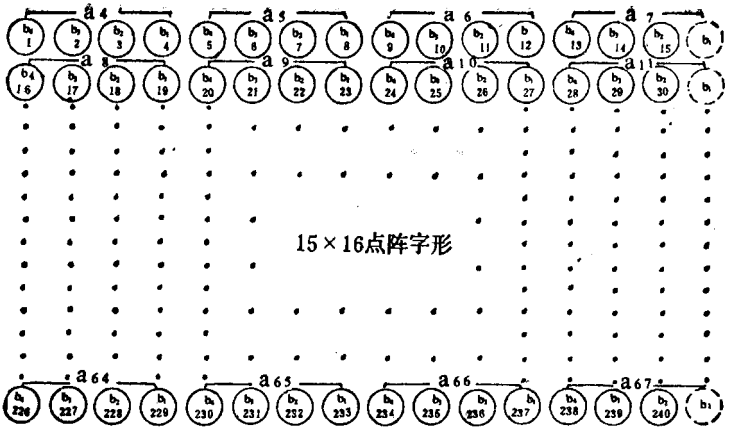


图 5·17　15×16点阵字形的构成和装配（非连续）

16进制数的最高位（b_4）装配。如图5·17中的第一扫描行，用a_4、a_5、a_6、a_7 4 个16进制数对应装配，其中a_7的最低 位 b_1 是空位，始终是“ 0 ”。第二扫描行则从a_8的b_4位开始装 配，行末的a_{11}最低位（b_1）也空着不用。最后，用第64个数 据 a_{67} 装配238～240个点位。这样，15×16点阵字形需用64个16进制数（32字节）进行装配，比连续装配多用 2 个字节。实际上，对于15×16点阵字形的存储和传送，都是以16×16点阵的32个字节计算的。所以一般都采用非连续装配方式，处理比较简便。只有在传送和装配数量很大的字形或某些复杂图形时，可以采用连续装配方式，以提高效率。

国标GB5261-85中规定的控制功能内容很多，它适用于各种不同用途的文字和符号成形设备，包括图文的显示、打印和印刷排版设备。对于汉字终端或通用的汉字微机系统来说，并不要求同时具备该标准规定的所有控制功能。一般只需选用其中的一部分功能已能满足使用要求。但是，当选用标准中已有

的控制功能时，其编码表示和参数定义应与标准的规定一致，以保证汉字设备和系统的兼容性。当选用标准中还没有规定的专用控制功能时，应使用标准中空着的位组来表示，或者按照GB2311的规则进行扩充。

5·3 汉字通信

实现汉字信息交换是汉字终端的基本功能之一。有两类信息交换方式：一类是人与机器之间的信息交换，汉字终端作为实现人——机信息交换设备的各项功能，在本书前四章中已经作了较详细的讨论；另一类是机器与机器之间的信息交换，就是把经过处理的信息（或是由人输入的信息）按照规定的代码格式发送到对方的同类设备或计算机，或者接收对方发来的代码信息，这类机——机之间的信息交换就是汉字通信。本章前两节讨论的汉字交换码、控制功能码就是为实现汉字通信而制订的统一数据格式。

目前国内汉字通信有两种方式，即电报通信方式和数据通信方式。汉字电报通信方式就是在现有电报通信系统的基础上，把适用于西文通信的字符式电传机改造为能自动接收和打印汉字电报格式的汉字电传机。通信系统的其它特性如电码制式、传输速率、信道设备、交换方式、通信规程等都保持不变，这样有利于现有电报通信设备的充分利用，尽快实现电报通信的汉字化。就大多数电报通信技术人员来讲，这是一种比较熟悉的汉字通信方式，故不赘述。本节主要介绍有关汉字数据通信方面的内容，数据通信涉及的范围很广，专著较多，本节从介绍数据通信的若干基本概念入手，简要讨论与汉字终端技术关系比较密切的几个问题。

1. 汉字数据通信概述

数据通信是既古老而又年轻的一门通信技术。早在电话发明之前，电报通信就作为一种远距离传输信息的手段得到了应用。现在，电报通信已发展成为在世界范围内传递书面信息的有效通信手段。现代数据通信是随着电子计算机的广泛应用而诞生、是以通信技术与电子计算机技术相结合为特征的新型通信技术。与电报通信一样，数据通信也是传递离散的数据（字符）信息，但是，数据通信超越了电报通信主要是终端与终端之间交换信息的范围，可以实现终端与计算机之间的通信。因此汉字终端采用数据通信方式，不但比电报通信方式有更高的传输速率和准确性，而且各种通信功能更加完善，使用更为方便。随着各类数据通信网的发展，汉字数据通信的应用必将更为广泛和普及。

（1）汉字数据通信链路的基本结构

汉字数据通信链路的基本结构如图5·18所示。它主要是由数据终端设备和数据传输设备两大部分构成的。数据终端设备是指汉字电传机、汉字终端设备以及带有通信接口的汉字微机等；数据传输设备主要指调制解调器和信道传输设备等。

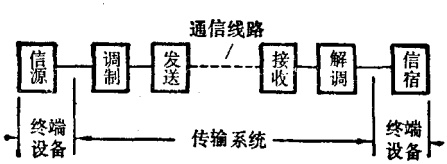


图 5·18 数据通信链路的基本构成

① 数据终端设备

数据终端设备简称为DTE，是发出或接受汉字信息的终点站，又称为信源或信宿。它的基本功能是把离散的汉字和字符数据变换成统一格式的数字信号。这个变换过程称为信源编码，以便向传输设备提供与传输特性相适应的代码格式和选择适当的传输速率。例如，对于同步传输系统来说，信源应发送规定时序和格式的同步电码信号，并可用较高的速率传送；对异步（起止式）传输系统来说，信源应发送起止式电码信号，传输速率一般不超过2400波特。为了提高传输信号的可靠性，在信源编码中应采取初步的差错控制措施，如在传输电码中增加奇偶校验码元和校验序列等。当需要进行保密的汉字数据通信时，在信源编码之后还需进行汉字数据加密处理，才将传输信号发送到数据传输系统去。

信宿解码是信源编码的逆过程，接收端把传输系统送来的数据信号以相反的过程还原成可以被信宿处理的内部码形式。经过处理后以文字或其它形式输出。

② 调制解调器

调制解调器是调制器和解调器的合称，有时也称为数(据)传（输）机。调制解调器是数据终端设备与信道传输设备之间的连接界面。在汉字数据通信中，通常以调制器或解调器来标志数据传输系统的起始或终止，故又称它为“数据电路终接设备”，简称DCE。

调制解调器的主要作用是将数字信号变换成能够在模拟电路上传输的形式。因为现有的绝大多数信道是用来传输连续变化信号的模拟信道，有一定的传输频带限制。由信源来的数字脉冲信号具有极宽的频带，通常不能直接在模拟信道上传输。因此，必须用调制器将数字信号“整形”，消除其中的某些高

频成分；并把有用的基本信号频率（基频）调制到较高频率的载波上，使之与所提供的信道带宽相匹配。例如，速率为100波特的汉字电传机信号，它的基本信号频率范围为0～50赫。如果用频带为300～3400赫的普通电话信道传输，则必须将这个0～50赫的基带信号搬移（即调制）到话音频带范围内；如果还要用无线信道传输，则需进一步调制到具有电磁辐射特性的高频或超高频频带范围内。

解调是调制的逆过程，解调器将接收的载波信号还原成频率较低的基带数字信号，并将它们传送给信宿。

调制解调器在改善和提高汉字数据系统的传输性能方面具有重要作用，通过各种复杂的调制方法，可以实现信道的多路复用，大大提高信道利用率，并能补偿各种信道干扰对传输质量的有害影响。

调制解调器的种类很多，用于在公用电话网上传输的普通调制解调器，可以为汉字数据终端提供300至600比特/秒的传输信道；质量较好的调制解调器可以在专线线路上把传输速率提高到9600至14400比特/秒，群路调制解调器则具有更高的信息传输速率。

应当指出，当直接用数字电路来取代模拟电路传输汉字数字信号时，调制解调器的许多功能将不再需要，可以用较简单的数字复接设备代替调制解调器作为“数据电路终接设备”(DCE)。

③ 数据传输设备

数据传输设备是连接信源和信宿的主体，是数据信息传输的媒介和通道，故又称为信道。

用于汉字数据通信的信道可分为有线和无线两类。有线信道如用电缆或光缆传输的大容量载波信道；无线信道如微波接

力或卫星信道。如前所述，速率较低的汉字数据信号通常只需占用公共电话网或专线的音频信道传输。

不论采用何种信道，在信号传输过程中总会产生衰减和畸变，还会受到各种无法预知的信号干扰。为了提高有用信号的功率电平，尽量减少衰减和干扰带来的不利影响，在传输信道的发收两端必须进行必要的信号放大和滤波，如有线载波设备中的放大器、滤波器及无线设备中的发射机、接收机等。但是这些设备并不对信号进行改变其特性的处理，仅起提供通路的作用，故常被看作是信道的一部分。

按照信息传输方式的不同，数据信道可分为单工、双工和半双工三种类型。

单工信道是只能提供一个传送方向的信息传输通道，参见图5·18。它只能从信源向信宿传送信息，不能反向传送，因此称为单工信道。

双工信道是能同时进行双方向传送的信息传输通道，通常采用频率分隔或时间分隔技术，将信道分成发送信道和接收信道，也可以用两对线路来构成双工信道。当使用双工信道进行汉字数据通信时，就需将发、收对应的各种设备装在一起，如图5·19所示。其中有包括信源和信宿的终端设备，包括调制器和解调器的线路终接设备以及包括发射机和接收机的信道传输设备等。有了这些设备，便可进行两端同时发、收的双工（又称全双工）汉字通信。

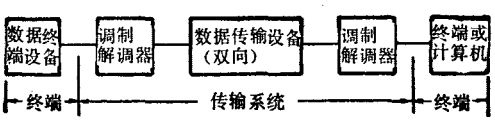


图 5·19　双工汉字数据通信链路

半双工信道可以提供双方向的信息传输通道，但不能同时进行传输。汉字数据通信中的半双工通信方式，就像打电话时必须等待对方讲完以后，你才能向对方讲话一样，两个端站不能同时发送（或接收）信息。半双工工作时，汉字数据通信链路中的设备配置与全双工一样，但传输信道可以节省一半。故在汉字数据业务量不大的场合大多采用半双工通信方式。在干线链路或业务量较大的场合下使用的汉字终端设备，则应采用全双工通信方式。

（2）汉字信号的传输方法

汉字数据通信的根本目的就是将代表汉字和字符数据的二进制数字信号由发方传送到收方。有两种数字信号的传输方法：并行传输和串行传输。并行传输是把一个字符代码中的各个码元，通过多条并行传输线同时传送。例如，对8位字符代码就用8条线传送，并行传输的速度极快，但要付出很高的线路设备代价，因而不适用于较远距离的通信系统，通常作为设备内部信号传输之用。在汉字数据通信中的数字信号都用串行传输方法。

串行传输是将汉字和字符代码的各个码元（“位”），在同一线路上以时间顺序一位一位地传送，接收方再将这些码元还原成相应的汉字和字符代码。

串行传输又分为同步传输和异步传输两种不同方式。采用同步传输方式时，汉字数据信号的每个码元都以固定的时间间隔T来发送（见图5·20），整个码流中任意两位码元之间的时间间隔均是T的整倍数。为了达到正确传输，发收双方的定时信号必须保持严格的一致，称为“同步”。采用异步传输时，同一字符代码中的码元也是以固定的时间间隔发送，但字符与

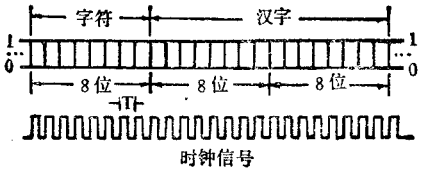


图 5·20 同步信号传输

字符之间的时间间隔是任意的。也就是说，在下一字符起始码元的开始时刻之前，始终保持在上一字符的停止码元状态，用这样的方法来保持字符同步，故异步传输又称为起止同步方式或起止式。图5·21所示是异步（起止式）传输方式的两种电码格式，图（a）是五单位起止式信号，图（b）是七单位起止式信号。

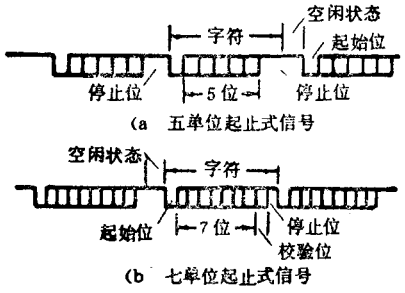


图 5·21 异步（起止式）传输

同步传输对发、收设备的定时精度要求严格，设备价格比较昂贵，适宜于进行中、高速数据传输；异步传输对设备的定时精度要求不高，成本较低，适合于进行低速数据传输，它不需单独的时钟信号，而用起止信号来保持字符同步的方法，尤其适用于进行会话式通信的各类汉字终端设备，如汉字电传

机、汉字显示终端等。

（3）汉字数据传输速率

传输速率是衡量汉字数据通信系统性能的主要指标之一，并与汉字终端的设计和使用密切相关。有三种常用的数据传输速率，即信号速率、信息速率和消息速率。

① 信号速率R_B

信号速率又称为调制速率，是在单位时间内传输信号的码元数目，单位是波特（Baud）。例如，在图5·22中，一个字符的起止式五单位电码传输时间为150毫秒，即每个码元的传输时间宽度为$\frac{150}{7.5}$=20毫秒，也就是相当于在一秒钟内传输50个码元，信号速率等于50波特。

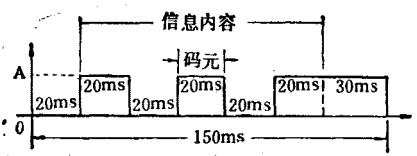


图 5·22 起止式二电平码元信号

应当注意，传输信号的码元可能是二进制、也可能是多进制的情况。就是信号码元除了“有、无”或“正、负”两种状态外，还可能出现不同电平值的多种状态，如图5·23中具有A_1、A_2和$-A_1$、$-A_2$四种电平值。这样，每个码元就可表示四种信号状态，即00、01、10、11，能携带较多信息量。但由于码元时间宽度不变，故信号速率仍为50波特。

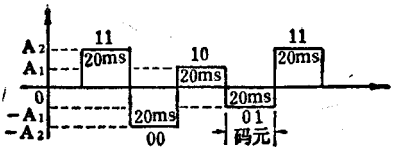


图 5·23 起止式四电平码元信号

② 信息速率R_b

信息速率又称为数据比特率，是指单位时间内所传输的信息量。

信息量是对信息多少的一种量度。因为各种信息最终都可以用二进制的位组来表示，一个二进制位就是一比特，所以实用中是以比特作为信息量的单位。信息速率等于一秒钟内传输的信息量，单位是比特/秒（bit/s）。

信息速率与信号速率在数值上的区别，取决于每一个信号码元所带的信息量。一个二进制码元的信息量为1比特，信息速率与信号速率在数值上相等；对于非二进制的码元信号，两者的数值是不等的。

设码元是M进制，则一个码元带的信息量为：

$$K=\log_2 M \text{（比特）}$$

若已知信号速率为R_B，信息速率R_b为：

$$R_b=K\cdot R_B \text{（比特/秒）}$$

以图5·23为例，图中所示信号速率为50波特，是四进制码元，$M=4$，$K=\log_2 4=2$，则信息速率为：$R_b=2R_B=100$比特/秒。

在图5·22中，因为码元是二进制，$M=2$，

$$K=\log_2 2=1$$

则：$R_b=R_B$

由此可见，在信号速率相同的条件下，采用多进制码元，

可以增加信息传输速率，有效地提高信道的利用率。这对充分利用线路的传输能力是非常有用的。但是，采用多进制码元增加了设备调制（编码）和解调（译码）过程的复杂性。

③ 消息速率

消息速率又称数据传送率，是指单位时间内传输的消息数目。CCITT给消息速率下的定义是："单位时间内，在数据传输系统的相应设备之间通过的每秒、分或小时中的平均比特、字符或码组数"。

例如，用起止式五单位电码传送字符，每个字符有7.5个码元，在信号速率为50波特时的消息速率为400字符/分。用起止式七单位电码传送时，每个字符有11个码元，当信号速率为110波特时，消息速率为600字符/分。我们可以用同样的方法计算传输汉字时的消息速率：当以汉字电报码的格式传输时（参见图5·4），若信号速率为50波特，则汉字消息速率为100字/分，如以字与字之间加上"间隔"字符计算，则消息速率为80字/分；当用汉字国标码的格式传输信号速率为110波特时，则汉字消息速率为300～200字/分（每个汉字以2～3个七单位字符代码计算）。

以上三种速率指标既有联系又有区别，在电报和低速数据的异步数据通信中，一般用传输信号速率(波特)为标准，如50波特的电传信号、300波特的低速数据信号等。在速率较高的异步或同步数据通信中，一般用传输信息速率(比特/秒)为标准，如1200比特/秒的异步数据信号或9600比特/秒的同步数据信号等。计算消息速率时，还应该包括两个端站的业务处理时间。

（4）差错控制和检错编码

在汉字数据通信系统中，由于脉冲噪声、瞬时中断以及其

它不可预知的干扰因素的影响，使传输的数字序列发生错误，如发送的是“0”信号而接收时变成了“1”信号，还有信号码元的丢失或增加等等。从错误发生的形式来分，大体有两种：一种是偶然性的互不相关的错误，称为随机错误。这种错误大多发生在微波接力信道、卫星转发信道或有线信道；另一种是成串、成片出现的比较密集的错误，错误之间有相关性，称为突发错误，这种错误常发生在无线短波信道和散射信道。上述这些错误，都会对通信造成不同程度的损害，而对汉字数据通信的影响尤为严重。如由于某种偶然干扰或瞬时中断引起的随机错误对电话通信的影响不大，有时可忽略不计，但对汉字数据通信有显著影响。有时一个码元的错误或丢失，会使字符代码出错，甚至造成数据传输的延迟或停顿。由此可见，控制差错是提高汉字数据通信可靠性的一个重要内容，而检（纠）错编码技术则是差错控制技术的基础和核心。

① 差错控制的基本方法

差错控制的基本方法是在发送的数字信号序列中，以一定的规则预先加入检验码元，使原来信息码元之间互不相关的数字序列变为具有某种相关性的序列。于是，可以在接收端按照已知的规则进行检验，如果检验结果不符合这种确定的相关性，则可知产生了传输错误。在检出错误以后，以纠正错误的不同方法区分，有以下三种差错控制方式：

（A）前向纠错（简称FEC），又称自动纠错，这种纠错方式的特点是接收端在收到信息后自动纠正传输中的错误。前向纠错方式使用比较方便，能用于实时单向通信，不需要反馈信道，发送的信息也不需储存。但前向纠错必须采用专门的编码方法，解码设备比较复杂；为了纠正较多的错误位，需在码组中携带的检验码位比较多，传输效率较低。

(B) 检错重发（简称ARQ），又称判决反馈或回询纠错，这种纠错方式的特点是接收端在收到信息后，检验传输中有无错误，然后把检验结果通过反馈信道送回发送端；发送端根据接收端的反馈信息把有错误的信息重发一次。如重发的信息接收端仍检验有错，则再反馈重发一次，直至接收端判定无错为止。

检错重发方式的具体实现方法有两种：一种是发送等候方式，发方发完一码组后就等候收方的反馈信息，若收方表示无错就继续往下发送；若有错则把已发码组再重发一次。另一种是连续工作方式，发方一组接一组连续不断地发送信息，同时不停地接收和检查收方送来的反馈信息，若有错误，发方就重发。重发的方式也有两种：一种是从有错的一组信息开始重发，称为拉回方式；另一种是只重发有错的一组，称为选择重发方式。不论采用何种重发方式，都要求接收方在反馈信息中指明错误在哪一组，同时要采用抗干扰能力强的特殊编码来保证重发指令的可靠传输。

检错重发方式的编码方法和编解码设备都比较简单。只需用较少的检错码元并与简单的传输控制规程结合起来，便可得到比前向纠错高得多的差错控制能力，因而在各种有线、无线数据通信中得到广泛的应用。但是，检错重发方式需用一条反馈信道，因此，在某些单工传输系统中只能使用前向纠错方式。

(C) 混合纠错（简称HEC），这是将前向纠错与检错重发结合使用的方式。发方发送同时具有检错和纠错能力的码组，收方收到后检验错误情况。如果错误码元数较少，不超过纠错能力，则自行纠正，不需反馈重发；如果超出纠错能力，则经过反馈要求发方重发。这种方式兼有前向纠错和检错重发的优点，又弥补了它们各自的不足，因而近年来有较多的应

用。

② 纠（检）错编码概述

从上述讨论可知，不论采用何种差错控制方式，都是以纠（检）错编码为基础。发送端发送具有纠（检）错能力的编码序列，接收端通过对这种特殊编码序列的检验，才能发现或纠正错误，达到控制差错的目的。

信息论的理论分析指出：如果数字信号传输速率R小于信道所能允许的最大传输速率C（称为信道容量），那末总存在一种编码方法，可使错误解码概率随码长n的增加而指数递减。这就是说，只要保持R＜C，通过增加码长n，总可以找到一种把错误控制在允许数值以下的编码方法。这是实现纠（检）错编码的理论基础。由此可见，用纠（检）错编码来提高汉字数据通信的可靠性是以牺牲通信的有效性为代价的。因此，应根据系统的使用要求和各种信道的不同特点，采取相应的纠(检)错码方法。不能简单地认为，在任何情况下都是差错率越小越好。

在汉字数据通信中常用的检错编码方法有以下几种：

（A）垂直奇偶校验

奇偶校验是最常用的一种检错编码法。奇偶校验有垂直奇偶校验和水平奇偶校验两种，通常所说的奇偶校验，如果没有特别说明，就是指垂直奇偶校验。

垂直奇偶校验的方法是在汉字或字符的七位代码之后，又附加一个第8位，即奇偶校验位。第8位的取值原则是，应使新的8位代码中的二进制“1”的个数始终保持为奇数（奇校验）或偶数（偶校验）。垂直奇偶校验的字符代码格式参看图5·1，图中b_8就是奇偶校验位。

假设有一字符代码为 1 1 0 1 1 0 1（小写字母m），

若采用垂直奇校验，则应在校验位置“0”，成为新的8位代码 0 1 1 0 1 1 0 1，使码组中“1”的个数仍为5个（保持奇数值），接收方通过对每一码组中“1”的个数进行检测，便可测知是否在传输中发生了错误。如上例中的b_7位由1变0，接收到的码组为 0 0 1 0 1 1 0 1，其中“1”的个数已变成偶数，便知发生了传输错误。显然，这种校验法不能检出偶数个码元的传输错误。

垂直奇偶校验法简便易行，已在汉字数据终端中得到普遍应用。通常在起止式的汉字数据终端设备中采用偶校验法（使码组中“1”的个数成偶数）；在同步式汉字数据终端设备中采用奇校验法（使码组中“1”的个数成奇数）。

（B）水平奇偶校验

垂直奇偶校验是以字节为单位的检错编码方法。为了进一步提高检错能力，还可以采用以字节串为单位的水平奇偶校验。

水平奇偶校验又称纵向奇偶校验。它是在保留字节垂直奇偶校验的基础上，将发送的字符代码分成若干码组（即字节串），在每个码组后面附加一个按照预定规则生成的码组校验字符（简称BCC），接收端根据相同的规则进行检验，便可发现错误。BCC的生成方法如图5·24所示。

由图可知，码组校验字符（BCC）也是由8位构成，它的b_1～b_7是码组校验位，b_8是BCC本身的奇偶校验位。b_1～b_7位取值的规则是取码组中各字符代码相应位的模2和。也就是说，使水平方向码元中“1”的总数（包括BCC本身在内）为偶数。如图中b_1位一行（最下一行）中原有5位是“1”，故BCC的b_1位为“1”；b_2位一行中原有4位是“1”，故BCC

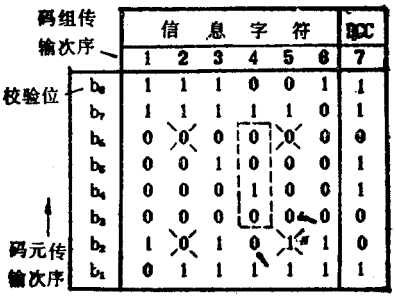


图 5·24 字符与码组奇偶校验

的b_2位为“0”。BCC的b_8位是自身b_1～b_7位的奇偶校验位，采用与系统相同的校验方法。如在同步传输中，BCC本身采用奇校验，图中BCC一列中已有4个“1”，故b_8取“1”。

接收端在收完整个码组以后，根据上述纵横两座标的奇偶校验关系，可以检测出所有长度$b \leqslant n$(n为码长，图中n = 8)的突发错误。比如图中虚线框内连续4个码元都发生错误，那末，b_3～b_6四个水平行的码元都不满足偶数校验关系，收方仍然能检测出来。可见，此法具有较强的检错能力。然而当发生图中X所示的矩形座标式错误时，仍然检测不出来。

（C）循环多项式校验

循环多项式校验又称循环码校验（简称为CRC），是一种更有效的差错检验方法。

CRC校验也是以码组为基础进行检验，发送端将需发送的字符代码分成若干码组，在码组后面附加一个按照预定规则生成的码组检验序列（简称BCS）。BCS通常由两个字节（16位）构成，接收端根据相同的规则进行检验，以便发现错误。

码组检验序列BCS是用代数运算方法生成的。就是在发送

端将待发送的码组乘以X^{16}，然后被一个多项式用模2算法相除，相除的余数形成BCS的校验位。一般采用的多项式是CCITT V.41建议（已由国标GB3453-82采用）的下列多项式：

$$P(X) = X^{16} + X^{12} + X^5 + 1$$

接收端将收到的信息码组和BCS除以多项式P(X)，若所得余数为零，则说明传输中没有错误；否则，便可检测出错误来。

循环多项式校验法具有很强的检错能力，而且可以通过严格的代数运算设计出各种实用的、具有更强纠错能力的循环码。因此，此法在实际工作中得到了非常广泛的应用。但是，随着检错能力的增强，也增加了编码的冗余度，故应在较好的突发错误检测能力与较低的冗余度之间进行权衡考虑，才能得到最佳的编码方案。

检（纠）错编码的理论和方法内容十分丰富。这里只是初步介绍与汉字通信关系较多的几种，有兴趣的读者可进一步参阅有关专著。

（5）汉字数据通信链路和网路的结构方式

在汉字数据通信中，把两个或两个以上端站（节点）连接起来，可构成具有各种不同性能特点的汉字数据通信链路和网路。现在简要讨论几种常用的链路和网路结构方式。

① 点对点方式

点对点的结构方式是汉字数据通信中应用最多、也是最基本的连接形式。图5·25所示是连接A站和B站的点对点汉字数据通信链路，DTE是汉字数据终端设备，如电传机、汉字终端、计算机等设备，DCE是调制解调器或数据复接设备，它们通常与终端设备放置在一起，但属于线路传输系统。连接两个端站的通信线路，可以是永久保持连接的专用线路，也可以

是在公共交换网中进行一次呼叫期间连接起来的公用线路。对于后者，汉字数据终端设备应该设有与自动电话相同的拨号呼叫装置。

Retro Workshop
果粉工作室

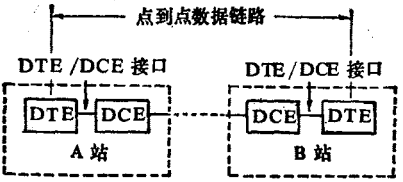


图 5·25　连接A、B站的点对点汉字数据通信链路

② 一点对多点方式

一点对多点的结构方式（简称多点方式）是由一个主站与若干个从站相连接的汉字数据通信系统，大体有链路形和星形两种形式，如图5·26所示，是A站（主站）和B、C、D三个从站的多点连接方式。图（a）是链路形连接，图（b）是星形连接。链路形连接时，A站用一条带有分支的线路，便可把B、C、D三个从站连接起来，这样比较经济，尤其当三个从站都在距离较远的同一地区时，可以节省大量线路费用。星形连接是各从站都直接连到主站，这在从站数量较多，而且距离很近时使用比较方便可靠。

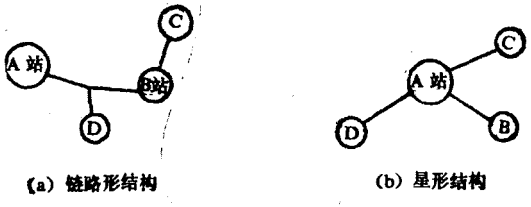


图 5·26　连结四个站的多点连接方式

在多点方式的汉字数据通信中，汉字信息传输只能在主站与从站之间进行，从站之间不能相互直接通信。例如，A站可向B、C、D三站轮流或者同时发送信息，也可分别接收来自B、C、D站的信息。为了防止各从站争用线路，或在A站接收信息时引起冲突，通常在A站采用一种叫做“轮询”(Polling)的通信规程。在电路空闲时，A站要定时轮流询问B、C、D站有无信息需要发送，如有信息传输，则须等待接收完了以后再询问下一站。

一点对多点方式是一种集中处理式的汉字数据通信结构，系统内各种控制功能都由主机实现，主要用于主计算机与终端之间的通信，也可以作为较大范围数据通信网的一个组成部分。多点方式的链路形结构，各站间应以专线连接；星形结构可以用专线或通过交换网连接。如果主机同时兼有交换功能，则多点链路结构就成为各从站之间也能进行直接通信的星形通信网路结构了。

③ 网路方式

网路结构方式适用于有多个主站和多个从站的汉字数据通信系统的连接。

随着汉字数据传输业务量的增加和传输范围的扩大，局限于两个端站之间或者主、从站之间进行通信的链路方式已难以满足需要，于是便有各种结构形式的汉字数据通信网路出现。数据通信网以计算机技术与通信技术的结合为特点，能通过较少的线路设备，将网路内的各个端站连接起来，实现直接的汉字数据通信。

与电话通信网路一样，数据通信网路也是由一组节点和连接节点的链路构成。节点的主要设备是交换机，用以支持网路的连续性，通过它来接通链路和传递信息。链路是节点间的信

息通道，可以采用不同的传输介质，如各种电缆、光缆、微波接力信道和通信卫星信道等。根据不同的使用要求，可以连接成各种不同的网路形式，图5·27所示是四种常见的数据网结构形式。

图（a）是环形网，网内各节点按照环状连接，各节点间的信息通过多路复用设备在环内按照一个方向传送。该网还可通过某个连接节点（如图中D节点）与其它环形网相连，便于扩展。环形网所需的线路总长度较少，适合于信息量不大的场合使用。

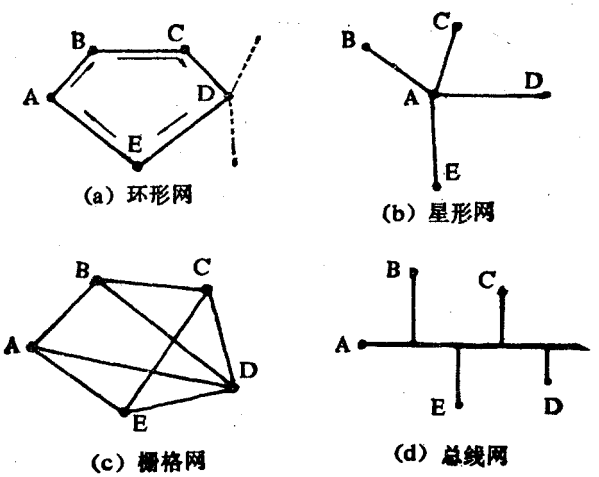


图 5·27　四种常见的数据网结构形式

图（b）是星形网，是由一个节点和若干端站构成的辐射状网路。它的结构形式与多点方式相似。图中A是一个节点交换机，主要是起各端站之间的连接作用，通过它能实现各端站

间的点对点通信或多点通信。星形网结构比较简单，所需线路费用较少，但安全性差。一旦节点发生故障就会使全网都不能工作。

图（c）是栅格网，是由多个节点组成的分布式栅格状网路。网中每个节点都可以通过链路方式与若干个端站相连接（图中未画出端站）。从图中看出，栅格状网路的每对节点之间至少有两条通路，任何一条通路中断，节点可自动寻找迂回路由接通。栅格状网的安全性好，便于网路的扩展和移动，适合于部队野战通信和要求可靠性高的其它场合使用。但栅格网所需的线路设备多，建设投资费用比较昂贵。

图（d）是总线网，是用一条公用总线把各个端站连接起来，它类似于计算机总线方式，适用于范围不大的局部网路使用。

（6）汉字数据通信网的交换方式

对现代通信系统的一个基本要求是系统内的任一用户可与任一其它用户相互直接通信，因此，一个有效的汉字通信网路必须具备交换能力。通过各种交换方式，把需要相互通信的用户接通链路；通信结束后，链路断开，可以为其它用户提供服务。这样不但提高了通信效率，还可以大大提高线路利用率，节省设备投资费用。

通信网的交换技术随着计算机技术的发展而日新月异。交换方式的变化又对通信网的结构、性能以至终端设备的功能产生重要影响。

汉字数据通信网的交换方式大体有电路交换、报文交换和分组交换三种：

① 电路交换方式

电路交换是通过用户（终端）的呼叫，交换机将主叫用户与被叫用户直接连通电路，我们熟悉的电话和用户电报通信都采用这种交换方式。在汉字数据通信中，对600波特以下的低速汉字数据终端也可通过公共电话网的电路交换方式实现连接，再通过调制解调器进行传输。图5·28就是利用电话网传输汉字数据的一种简便方法。

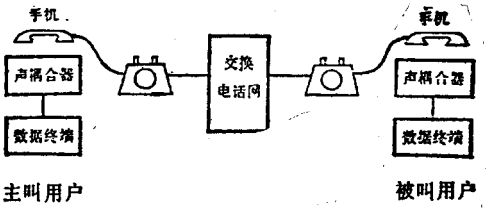


图 5·28　通过电话网进行汉字数据传输的方法

图中的声耦合器实质上是装有声电转换器（送话器和受话器）的调制解调器，调制方式是移频键控（FSK）。声耦合器的一端通过标准接口与汉字数据终端相接，并将终端来的数据信号转换为可听见的高、低频单音，然后馈送入电话机的受话器内；对方的声耦合器把来自电话机的可听单音转换为二进制数据信号，经过接口电路传送给汉字数据终端。用图5·28所示的方法进行低速汉字数据通信，可不必改动任何接线，只需用电话机把对方呼叫接通以后，把电话机的送话器和受话器分别扣在声耦合器的受话器和送话器上就行了。

电路交换方式的通信规则比较简单。两个用户之间一旦实现电路连接，就可方便地进行不受时间限制的实时会话式汉字通信。双方的汉字数据终端设备只要保持码型和速率一致，对通信内容和文本格式一般不作特殊要求。但是，由于电路交换方式在通信过程中始终占用一条电路，故电路利用率比较低。而且

一般情况下呼叫接续过程比较费时，当业务繁忙时容易产生电路阻塞，影响汉字数据通信的及时性。

② 报文交换方式

报文交换方式是一种存储转发交换方式，报文交换机是由存储器和转发器等单元组成。它并不进行接通或断开电路的实际操作，只要入口交换机的存储容量足够大，汉字数据终端可以在任何时候将报文发送到交换机储存起来，排队等候转发。当排队轮到时，就将报文转发到下一个交换机，这样重复存储转发，直到报文被传送到目的地为止。一旦报文通过某一个电路以后，不同目的地的其它报文就可再使用这个电路，因而可以提高电路利用率。但是，由于报文经过数次存储转发，需有一段延迟时间才能传到对方，有时时延可能较长，故不宜于进行实时的会话式汉字通信。

为了使交换机能准确地将报文传送到接收方，在报文交换方式工作时，发送端在所发送的报文中必须附加一些辅助信息，如报文的源地址、目的地址、优先等级、差错控制序列和其它传输控制信息。在特定的交换网中，这些附加信息必须以固定的标准格式加入报文的前后，作为“报头”或“报尾”与正文一起传送。图5·29是报文交换方式的一种传输信息格式的举例。

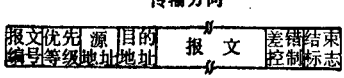


图 5·29 报文交换格式举例

③ 分组交换方式

分组交换是存储转发交换的一种新形式，它是在报文交换的基础上进一步发展起来的，是特别适用于计算机汉字数据通

信的一种交换方式。

分组交换与报文交换之间的主要区别在于：报文交换时，由于传输报文长度不受限制，任何一份长度不等的报文都在入口交换机接收、存储、转发，直至传送到接收端。网路对长报文的多次存储转发，必然使其它报文的排队等候时间延长，无法进行实时的会话式通信，有时甚至可能产生阻塞。在分组交换中，发送端把报文（数据）分为固定的、长度较短的分组，在网路中进行存储转发时，由于分组短、长度固定，增加了路由和流量控制的灵活性，大大减少了排队等候时间，使信息传输的时延比报文交换时短得多，而信息通过量却比较大，可用于进行会话式汉字数据通信。此外，分组交换还便于实现信道的时分复用，可用于包括数字话音信息和图象信息在内的各种信息业务的综合通信，是一种具有广阔应用前景的数据交换方式。

与报文交换一样，分组交换的传输数据分组必须具有固定的格式。图5·30是一种分组交换格式的实例。

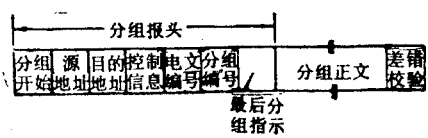


图 5·30 分组交换格式举例

由图可见，分组交换格式与报文交换格式不同之处是在报头中增加了“分组指示”的内容。由于分组的长度是固定的，故“分组结束”的内容可以省略。

分组交换有两种工作方式：数据报方式和虚呼叫方式。

数据报方式的工作过程与报文交换方式较为相似，但发送端需把一份完整的报文分成符合格式要求的若干分组数据送入网路中去。为了保持每个分组的长度相同，最后一个分组往往

需要增补一个特定码型的填充段。当网路把各个分组传送到目的地址以后，接收端对分组数据进行差错校验，并核对接收地址无误以后，去掉各分组的报头、报尾和填充段，重新组合成原来的完整报文。各个分组依靠报头指示信息在网路中独立地进行处理和传送，彼此互不相关。

虚呼叫方式是在分组交换的用户双方进行一次通信之前，先初步交换一些分组，类似于电话网中的呼叫建立过程。呼叫一旦被接受，双方建立起临时的逻辑信道连接，即被赋予一个逻辑信道号，随后传送的全部分组都带有这个逻辑信道号，在规定的物理电路上传输，就象双方始终占用了一条电路一样。但是，实际上只有当上述的分组数据传送时，才实现物理电路和设备的实际连接，故称为“虚电路”。为建立“虚电路”的呼叫称为“虚呼叫”，在一条物理电路上可以建立若干条虚电路，一个用户也可以通过若干个虚呼叫同时与几个其它用户通信。

分组交换的数据报方式不需呼叫建立过程，接续快，传输延迟小；网路只处理单个分组电文，管理功能和接口协议都比较简单；适用于通信量小和要求传输时延小的场合。虚呼叫方式采用电路的统计复用方法，电路利用率高，使用方便灵活；由于实现了分组顺序和流量控制，传输可靠性很高；但网路管理功能和接口协议相当复杂，设备价格昂贵；适合于通信量大、对通信可靠性要求高的场合使用。

2. 汉字数据通信的接口和规程

在汉字数据通信系统中，为了使各种汉字终端或汉字计算机系统通过传输设备能协调一致地进行信息交换，就需要有统一的接口和传输规程。例如，在汉字电报通信中，电传机同外

线怎样连接，使用什么样的电码制式、报文格式和通报规则等，都属于通信接口和传输规程的范畴。显然，如果接口或规程不统一，或者机器与外线接不通，或者接通了也不能实现正常的信息交换。

汉字数据通信的接口和规程比电报通信复杂。由于数据通信实质上是计算机通信，它的终端设备、交换设备及网内控制设备基本上都是计算机，传输的数据量大，数据处理的自动化程度高，在通信过程中过去需要由人来判断处理的功能大多由计算机自动完成。特别是当把不同结构和性能的计算机连接到同一个网上时，情况就更为复杂，因此需要制订既能适应各种设备的不同情况，又应具备统一标准和兼容性的通信接口和规程。

为此，CCITT提出了如图5·31所示的四级规程的建议，图中自下而上分成物理级、链路级、分组级和用户级，现简要说明如下。

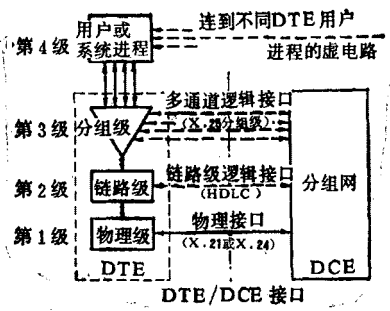


图 5·31　CCITT的四级规程

第1级是物理级。它规定了数据终端设备（DTE）和数据电路终接设备（DCE）之间物理线路的电气特性，被定为CCI

TT的X·21或X·24建议。

第2级是链路级。它规定了在DTE和DCE之间无差错地传输数据的必要条件和措施，使用国际标准化组织的ISO1745规程或HDLC规程。

第3级是分组级。它的基本内容是规定分组格式，并在需要时对各种信道送来的分组进行交织，以实现在单一的物理链路上复用多条逻辑信道。本级规程使用CCITT X·25建议的第3级（最高级）。上述的物理级和链路级规程则同时作为X·25建议的第1级和第2级。

第4级是用户级。它规定了各种数据终端和计算机之间通过分组交换网或固定虚电路建立联系、交换信息、调用文件和数据库等的用户应用规程。这些规程涉及所用不同计算机的硬件结构和操作系统，CCITT对此尚未提出具体建议。

随着分布式计算机信息处理系统的大量发展，使各种终端与系统、系统与系统之间进行通信以共享资源的要求愈来愈迫切。为了使它们的相互通信不受特定的设备和通信环境的影响，国际标准化组织在1978年提出了一个规程："开放式系统互连（OSI）的参考模式"。使凡是遵守这一规程的各个系统均可实现互连而彼此开放，成为一个开放系统，开放系统互连（OSI）的参考模型如图5·32所示。

OSI模型把一个通信终端系统的功能按逻辑结构划分成七层，并定义了每层的功能，其内容大致如下：

第一层物理层，规定DTE和DCE之间的物理连接和电气特性。

第二层数据链路层，它的主要功能就是实现基本型传输控制规程或HDLC规程。

第三层网路层，只适用于分组交换网，其主要功能是数据

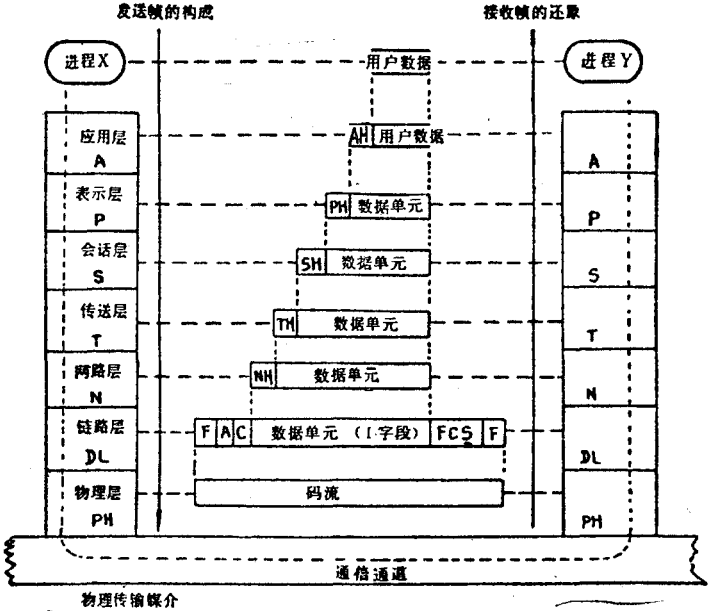


图 5·32 开放系统互连 (OSI) 参考模型

交换、路由选择、通信流量控制等。

第四层传送层，主要用于实现用户的“端—端”之间或进程之间的信息控制和信息交换。

第五层会话层，其主要功能是建立并保持用户进程之间的逻辑关系及处理进程之间的对话。

第六层表示层，主要用于对不同的数据结构和程序语句句法的变换，以实现异种机联网。

第七层应用层，主要是为用户进程进入开放系统提供相互连通的条件，实现把单机的用户功能扩大为网路的功能。

OSI参考模式的各级规程是按照用途和功能的不同来划分的。前3层（第1～3层）主要用于通信方面，其内容与CCITT建议的内容是一致的；后3层（第5～7层）主要涉及计算机应用方面；第4层则是通信与计算机应用的结合层。

OSI七层参考模式的制订，不但为数据通信网的建立提供了统一标准，而且为实现网间互连创造了条件，使甲网的任一用户终端能够与乙网的任一用户终端通信。当然这里还要解决大量的具体技术问题，OSI参考模式4～7层的具体建议尚在拟订中，已经拟订的许多规定还需不断充实和完善。本节简要介绍与汉字数据通信关系比较密切的前3层规程的主要内容。

（1）汉字数据通信的物理接口

接口是两个不同设备之间的连接端口，在数据通信系统中，通常是指数据终端设备（DTE）和数据电路终接设备（DCE）的连接端口或界面。为了能使各种类型的数据终端设备实现与数据传输设备的正确连接，接口的标准化是十分必要的。要使汉字终端进入数据通信系统进行汉字通信，也必须遵守DTE/DCE接口标准的有关规定。

如前所述，现有的传输信道绝大多数是模拟信道，为了能够传输数据数字信号，需要用调制解调器来进行模拟信号与数字信号的变换。因此，在数据通信中，通常以调制解调器来标志数据传输系统的起始和终止（参看图5·25），调制解调器就成为一种最常见的数据电路终接设备（DCE）。

但是，当用数字电路取代模拟电路传输数据时，调制解调器的许多功能将不再需要，就可以用较简单的数据复接器或集中器来作为数据电路终接设备（DCE）了。

汉字终端或其它数据终端设备（DTE），通过DCE接到

数据传输系统，其接口称为DTE/DCE接口。CCITT的V系列和X系列建议规定了DTE/DCE接口的电气特性和通过接口传送信号的具体控制功能。V系列建议是通过模拟电话网的数据传输规程；X系列建议是通过数字数据网的数据传输规程。除了电气特性和信号线的控制功能外，有关接插件的结构尺寸和插针分配的建议和标准，则由ISO的有关标准（ISO 2110或ISO 4902）规定。

① 汉字终端与调制解调器接口——CCITT V·24 建议

汉字终端与调制解调器接口标准主要是CCITT的V·24 和V·28建议。V·24建议规定了DTE/DCE接口电路的功能定义，V·28则规定了它的电气特性。下面简要介绍V·24建议的一些主要接口功能。

V·24建议的接口电路共有40多条，在汉字终端设备中常用的有 9 条。它们的编号及名称如表5·11所示。

由表5·11可知，接口电路 9 条线中，除信号地线（102）外，其余 8 条接口线有 5 条是调制解调器（DCE）至汉字数据终端（DTE）方向，有 3 条是汉字数据终端至调制解调器方向。其中103线和104线是汉字数据终端的数据发送线和接收线，其它几条接口线都是为发收数据做准备或者检测线路状态的操作控制信号线。它们的具体操作过程如图5·33所示。

当需在专线或公共电话网上建立一条汉字通信电路时，首先按照正常通话过程接通话路，双方沟通联系后把电话线路倒向调制解调器及汉字数据终端位置，便构成一条汉字数据通信电路，如图5·33上部所示。

这时，主叫方通过调制解调器向被叫汉字终端送入 125 信号，表示进入呼叫状态。双方的汉字终端分别向各自的调制解调器送入108/1或108/2电路信号，表示汉字终端已经做好准备，

表5·11　　V.24建议的主要接口电路

V.24电路编号	接口电路名称	传送方向		RS-232C电路编号
		调制解调器	汉字数据终端	
102	信号地或公共回路			7
103	发送数据	←	—	2
104	接收数据	—	→	3
105	请求发送	←	—	4
106	允许发送	—	→	5
107	调制解调器准备好	—	→	6
108/1	把调制解调器接至线路	←	—	20
或108/2	数据终端准备好	←	—	20
109	载波检测	—	→	8
125	进入呼叫	—	→	

把调制解调器接入线路；这时主叫方调制解调器则以107电路信号回答，表示已经作好准备。被叫方汉字终端以105电路信号表示“请求发送”，调制解调器以106电路信号回答“允许发送”，主叫调制解调器以109电路表示线路情况正常，于是被叫汉字终端就可以向主叫终端发送汉字数据了。

数据发送完毕，被叫汉字终端在105电路上送出“请求发送断”信号；主叫汉字终端在109电路上收到“载波检测断”，便可以相反的过程向被叫终端发送汉字数据了。

V·24建议作为通过模拟电路的数据通信接口标准，已经得到广泛的应用。我国的国家标准GB3454-82《数据终端设备（DTE）和数据电路终接设备（DCE）之间的接口电路定义表》，是以CCITT V·24建议完全等效采用的。人们熟知的RS-232C接口，是美国电子工业协会的建议标准（C是现行的修订版），其接口电路的功能特性和V·24也是相同的，只是接口电路的编号不同，RS-232C与V·24接口电路的相应编号对

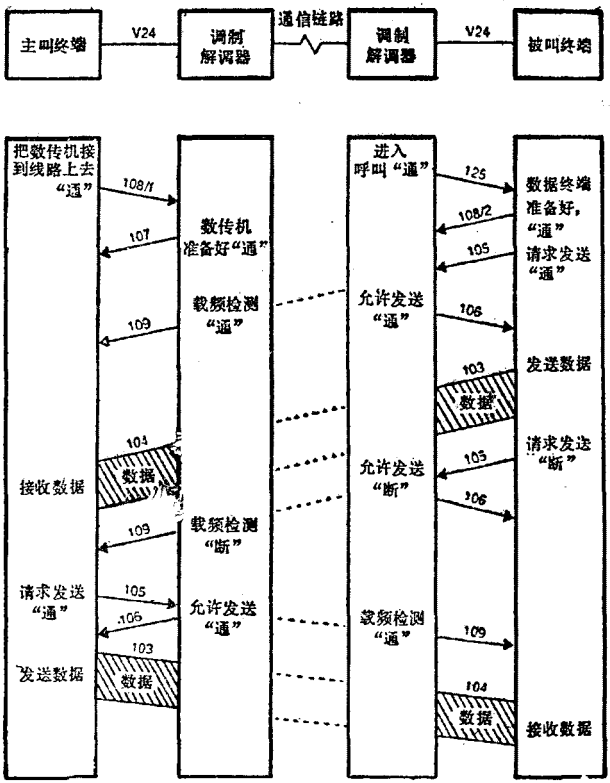


图 5·33 V·24接口的操作过程

照见表5·11的右侧一列所示。

② 汉字终端与数字数据网接口

——CCITT X·20 X·21建议

X系列建议是国际电报电话咨询委员会（CCITT）为适应公共数字数据网的发展提出的，建议包括的内容十分广泛，如终端复接方法、链路控制规程、数据传输业务规定等。而且随

着各种数据通信业务和通信方式的不断发展，X 系列建议的内容也在不断增加。这里介绍与汉字终端关系比较密切的X·20和X·21建议。

CCITT X·20和X·21建议规定了数据终端进入数字数据网的两种基本接口标准，当汉字终端需要直接进入数字数据电路进行汉字数据通信时，也适用这两个标准。其中，X·20 建议适用于起止式（异步）的汉字数据终端，可进行速率为 300 比特/秒以下的全双工传输；X·21建议适用于同步的汉字数据终端，可进行速率达48千比特/秒的全双工传输。X·20和X·21建议的接口电路功能如表5·12所示。

表5·12　　X.20和X.21建议的接口电路

电路标号	名　称	方向(至)	X.20	X.21
G	信号地或公共回路		●	●
T	发　送	DCE	●	●
R	接　收	DTE	●	●
C	控　制	DCE		●
I	指　示	DTE		●
S	信号码元定时	DTE		●
B	字节定时电缆屏蔽	DTE		(●)

由表5·12可知，汉字数据终端与数字数据网的DTE/DCE接口电路比较简单。如X·20的接口电路只需发、收、地三条线。这是因为许多功能是在数字电路的链路级或网路级完成了。数据电路终接设备（DCE）只作为某种形式的编译码器，将数据终端（DTE）送来的信号变换成适宜于线路传输的形式，而不再需要更多的附加控制功能电路。

（2）汉字数据通信链路传输控制规程

如上所述，汉字数据终端通过对DTE/DCE接口电路的连接和操作，在两个端站之间建立了数据通信链路以后，为了能够有效而可靠地进行通信，通信双方需要建立必要的规程或协议。例如对字符编码、传输控制功能、差错检测、文电格式及数据传送规则等内容，必须有统一明确的规定，否则，数据通信链路建立了也无法进行通信。但是，最初发展的一些通信规程往往只适用于某一专用系统，相互之间不能兼容。随着数据通信的发展和较大范围数据通信网的建立，国际标准化组织（ISO）在1971年发表了国际标准ISO 1745《数据通信系统使用的基本型控制规程》，后来又经过修改和补充，成为各国遵循的依据。我国的国家标准GB3453-82《数据通信基本型控制规程》就是参考上述标准以及其它有关标准制订的。基本型控制规程是被广泛使用的一种链路控制规程。70年代末以来，随着数据通信的迅速发展，基本型规程已不能完全满足要求，于是就有了数据通信高级数据链路控制规程，即ISO 4335-1984、ISO7809-1984等一系列标准。我国的国家标准GB 7575-87《HDLC规程要素汇编》、GB 7421-87《HDLC规程类别汇编》、GB 7496-87《HDLC帧结构》等标准，就是等效采用了上述国际标准制订的。

现将这两种数据通信规程中与汉字数据通信联系密切的有关内容简述如下：

① 数据通信的基本型控制规程

数据通信的基本型控制规程是一种面向字符的、以传输控制字符为基础的链路控制规程。它的主要作用是规定在已经建立物理电路的两个或多个终端之间，进行点对点或一点对多点

信息传送的控制规则。

基本型控制规程的主要控制功能包括：信息分组控制、传输差错控制、字符同步控制、信息流量控制，以及数据链路的管理、异常情况的恢复处理等。在扩充型的基本型控制规程中，还进一步提供了编码独立传输控制、双向同时传输等功能。

基本型控制规程的主要特点是：

• 采用标准的七位代码；以半双工方式进行通信；可以异步或同步传输；传输速率一般在48K比特/秒以下。

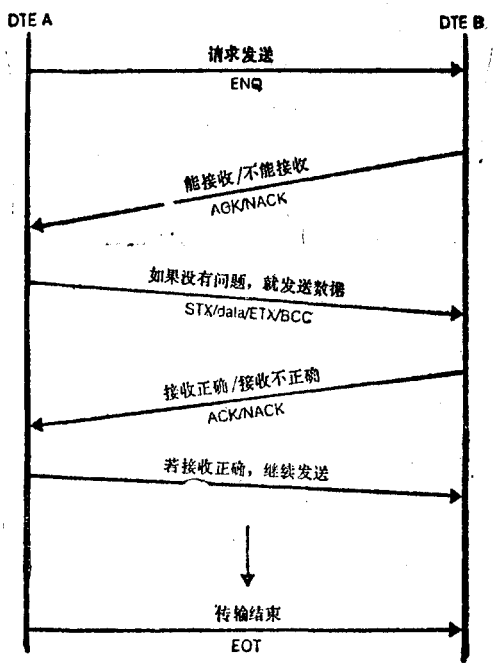


图 5·34 基本型规程的点到点操作

• 差错控制采用垂直/水平奇偶校验或循环冗余校验，用等待重发（ARQ）方式重发有错报文。

• 用GB1988七位标准代码表中的10个传输控制字符(TC_1～TC_{10}）实现全部传输控制功能。

现以图5-34来说明若干常用的传输控制字符的使用方法（参看5·2节）。

如图所示，DTE A（A站数据终端设备）首先用ENQ（询问）字符，即TC_5向对方表示要求发送数据，并请求B站数据终端设备DTE B作出应答；这时DTE B应以ACK（能接收）/NAK（不能接收）来答复对方。若答复ACK，DTE A就可以开始发送数据或报文，但必须在数据之前置以STX（正文开始）字符，即TC_2；在数据后缀以ETX（正文结束）字符，即TC_3；并在最后发送码组校验字符BCC。DTE A发送完毕以后，DTE B应以ACK或NAK字符，即TC_6或TC_8来表示接收是否正确无误。若接收正确，DTE B以ACK字符应答，DTE A就继续发送；若接收有错，DTE B以NAK字符应答，DTE A则应重发刚才发送的报文。

如此往复传输，一直到数据报文发送接收完毕，DTE A发送EOT（传输结束）字符，即TC_4，便结束了这一次传输过程。

• 基本型规程的报文格式有五种，大体可分成两类：有报头的报文和无报头的报文，如图5·35所示。图中BCC（码组校

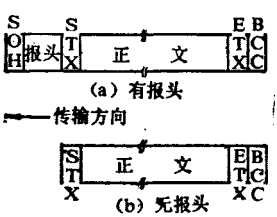


图 5·35 基本型规程的报文格式

验字符）是可选的。

在传输较长的报文时，可分成若干码组传送。但每个码组都以ETB字符结束，下一码组以SOH或STX开始。最后一组以ETX结束，表示整个报文已经发完。

·在扩充的会话型规程中，允许在ETX之后，对方可用SOH或STX来代替ACK作为肯定应答，这样可以缩短双方“交谈”时的转换时间。

基本型规程在接通线路后只能轮流单向传输，在传输的正文中不允许出现10个传输控制字符的代码码型，使传输的透明性受到影响。因此，它只适用于较简单的汉字数据通信系统。

② 高级数据链路控制规程

高级数据链路控制（HDLC）规程是一种面向比特的、以帧结构方式为基础的数据通信链路控制规程，它的主要作用是把易出错的物理电路改造成相对无差错的逻辑链路，以便在两个数据终端之间、或终端与网路之间实现高速可靠的数据传输。

与基本型传输控制规程相比，HDLC规程不但传输效率高、可靠性好，而且具有很强的传输透明性。也就是说，它可以传输任何一种格式或码型的数据代码，而不象基本型规程对传输代码格式（如7位或8位代码）和码型（如10个传输控制代码）有所限制，因而使用更方便。

HDLC规程之所以有不同于基本型规程的显著特点，主要原因是采用了以统一的帧格式来传输比特序列的新方法。在HDLC规程中，无论是报文信息还是控制信息，都看作是任意的一组比特序列，可以把它们组成统一格式的“帧”来进行传输。“帧”是HDLC规程中的传输单位，它的格式如图5·36所示。

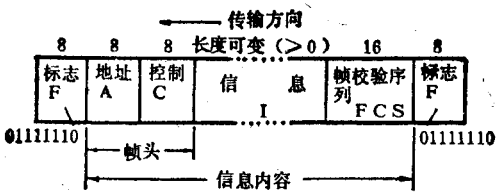


图 5·36 HDLC的帧格式

图中标志段F是一个固定的 8 位比特组 $\boxed{01111110}$，用来表示一个帧的开始和结束，这是通信双方建立帧同步所必须的。在连续发送帧时，同一个标志段既用于表示前一个帧的结束，又表示后一个帧的开始。当帧与帧之间出现间隙时，它还用作时间填空。为了保证标志段 $\boxed{01111110}$ 码型的唯一性，防止在信息比特流中出现同样的码型，规程采取了“0比特插入和删除”技术，规定在发送信息的比特序列中出现连续 5 个“1”之后，必须自动插入一个“0”，然后在接收端加以删除。

地址段A是一个 8 位比特组，用来表示该帧的站地址：如果发送“命令帧”，A为接收此帧的次站地址；如果发送“响应帧”，A为发出此帧的本站地址。

控制段C也是一个 8 位比特组，包含命令、响应及帧序号信息。“命令”是用于指示对方站完成某种操作；“响应”是对命令的应答。存在三种不同的控制段格式，分别为三种不同的帧，称为I帧、S帧、U帧，如图5·37所示。

由图可知，控制段C的第 1 比特为“0”是传送信息格式，称为“I帧”；C的第 1、2 比特为“10”是监视格式，称为“S帧”；C的第 1、2 比特为“11”是无编号格式，称为

控制段格式	控制段的比特位							
	1	2	3	4	5	6	7	8
信息传送格式（I帧）	0	N(S)			P/F	N(R)		
监视格式（S帧）	1	0	S		P/F	N(R)		
无编号格式（U帧）	1	1	M		P/F	M		

图 5·37　控制段的三种格式

“U帧”。I帧C段的第2～4比特N(S)是发送信息帧的序号，第6～8比特N(R)是等待接收的帧序号，即表示对已收到的N(R)-1个信息帧作出肯定应答。C段第5比特是P/F（探询/终了）比特，用于命令帧和响应帧的功能表示。

S帧主要用于链路状态的监视控制，例如表示肯定应答、探询、暂时中断信息的传递及差错恢复等功能，用C段的第3、4比特表示。S帧中不包含信息段，因此不需要发送N(S)，但需要发送等待接收下一个帧的序号N(R)。

U帧主要用于扩充链路控制功能的种类，例如建立与拆除链路、确定站的工作状态、探询、交换识别标记等。以上各种命令与响应均用C段中的5个M比特来表示，可以得到32种扩充的命令和响应功能。在发送U帧时不改变N(S)和N(R)值，因此称为无编号格式帧。

图5·36中，信息段I用于存放报文信息，它可以是若干比特的任意组合，与字符代码的格式无关，其长度也是任意的。

帧校验序列段FCS是16位的循环冗余校验（CRC）序列，构成对该帧（F段除外）的循环校验，其生成多项式为$X^{16}+X^{12}+X^5+1$。由于对不论报文信息或控制信息都进行校验，并且采用连续发送方式，即发方不等收方确认可以继续发送随后的数据。因此提高了通信效率和可靠性。

综上所述，高级数据链路控制规程能够在数字信道上以全双工工作方式传输各种二进制数据代码信号，并有很高的传输速率和可靠性。它不但可以传输各种汉字数据，而且也能传输数字电话和传真图形信号，已在各种类型的数据通信网中得到广泛的应用。

（3）数据通信网路分组级控制规程

数据通信网路分组级控制规程是CCITT X.25建议的第3级——分组级规程。它的主要功能是在物理级、链路级规程的基础上，为实现分组数据交换指定数据终端和网内设备的逻辑通道号，建立和拆除虚电路。所谓虚电路是指两个数据终端之间经过每次呼叫接通后建立的双向联系，这种联系是由分组经过的每一节点内的路由表建立起来的。数据信息通过这种双向联系进行交换，使两个用户之间好象存在一条数据通信链路一样。通过虚电路的建立，分组级规程将链路级规程提供的单一逻辑信道以统计复用方法变换成多个逻辑信道，并能独立地控制每个信道的流量。为实现上述功能，分组级规程对传送的数据分组（信息）和各种控制分组（信令），根据它们执行的不同功能，赋予相应的报头（标题）格式规定。现将X.25建议的分组类型名称列于表5·13。

表中所列的每一种分组都是用分组的不同格式和标识符来区分的，系统通过对分组格式和标识符的判别来执行各种操作和正确传送数据。现列举"数据分组"的格式及其在链路HDLC帧结构中的位置如图5·38所示。

由图可见，一个数据分组是由分组头和用户数据组成，占用HDLC帧的I段传输。分组头用3个字节，第1字节高4位是通用格式标识符，低4位是逻辑信道组号；第2字节是8位

表5·13　　X.25的分组类型和名称

分组类型	分组流向和名称	
	DCE⟶DTE	DTE⟶DCE
呼叫建立与释放	入呼叫 呼叫接通 释放指示 DCE释放确认	呼叫请求 呼叫接受 释放请求 DTE释放确认
数据与中断	DCE数据 DCE中断 DCE中断确认	DTE数据 DTE中断 DTE中断确认
流量控制	DCE接收准备好(RR) DCE接收未准备好(RNR)	DTE接收准备好(RR) DTE接收未准备好(RNR) DTE拒绝(REJ)
复　位	复位指示 DCE复位确认	DTE复位确认
重新开始	重新开始指示 DCE重新开始确认	重新开始请求 DTE重新开始确认
诊　断	诊　断	

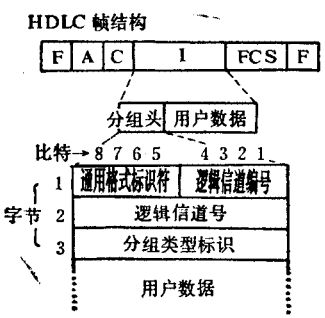


图 5·38　数据分组格式与HDLC帧结构

逻辑信道号，最多可表示256个逻辑信道；第3字节是分组类型的标识符。由于分组类型及其标识符的数量很多，限于篇幅这里不再一一列举了。

现在简要介绍通过分组网的两个数据终端（DTE A和DTE B）的通信过程。其中包括呼叫建立、数据传送、释放三个阶段，如图5·39所示。

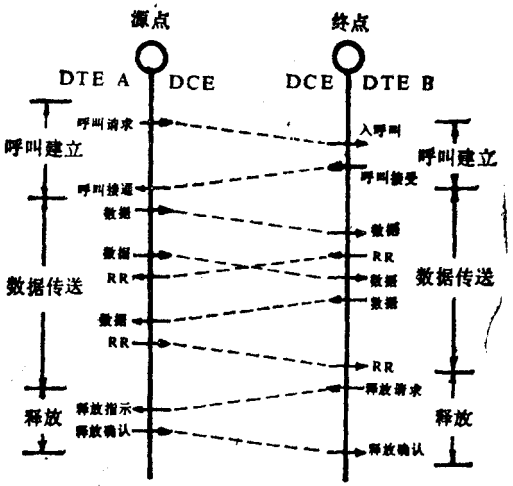


图 5·39　分组数据交换的通信过程

首先由DTE A发送“呼叫请求分组”到网路（DCE），表示它希望对DTE B建立虚呼叫，源点的DCE A在收到DCE B“呼叫接受分组”后，通过“呼叫接通分组”通知DTE A呼叫已经建立。进入数据传送阶段，DTE A连发两个“数据分组”。终点DTE B在接收第1个“数据分组”后，当时无数据可发，故发送一个“RR”分组，确认第1个“数据分组”

已经收好；在接收第2个“数据分组”后，DTE B若需发送“数据分组”，其中已包含了对第2个数据分组的确认信息，故不必另发“RR”分组。DTE A接收好“数据分组”后，如无数据发送，也发送“RR”分组。无论何方发起呼叫，一旦虚电路建立，双方都可以发送数据，通信结束时任何一方都可以发送“释放请求分组”。图中由DTE B发出释放请求后，DTE A收到源点DCE发来的“释放指示分组”，并发回“释放确认分组”，通信过程结束。

以上简单介绍了数据通信网分组级规程的基本内容及工作过程。对于一般只具备物理接口和链路传输控制规程的汉字终端和汉字微机系统来讲，要进入分组数据网，就必须通过分组组装/拆卸设备PAD。PAD的基本功能是把不具备分组级规程的汉字数据终端送来的字符组装成分组发往分组交换网，并把网路送来的分组数据中的用户数据部分取出拆成字符发给终端，从而实现两终端之间通过分组数据网的通信。终端与PAD之间的接口标准是CCITT的X.28建议。通过PAD，可使目前大量使用的汉字终端、汉字电传机等数据终端设备方便地接入分组数据网，对于今后研制的智能化汉字终端设备和汉字微机系统，则应同时具备X.25建议的三级接口规程，以便能适应未来数据通信发展的需要。

六、汉字终端的软件技术

汉字终端实质上是一个具有汉字输入输出和汉字通信功能的专用微型计算机系统，微型计算机是近十多年来发展最迅速、影响最广的现代高科技产品之一。随着微机技术日益广泛的普及应用，促使各类专用型的技术设备也发生了日新月异的革命性变化。汉字终端就是其中有代表性的一种。

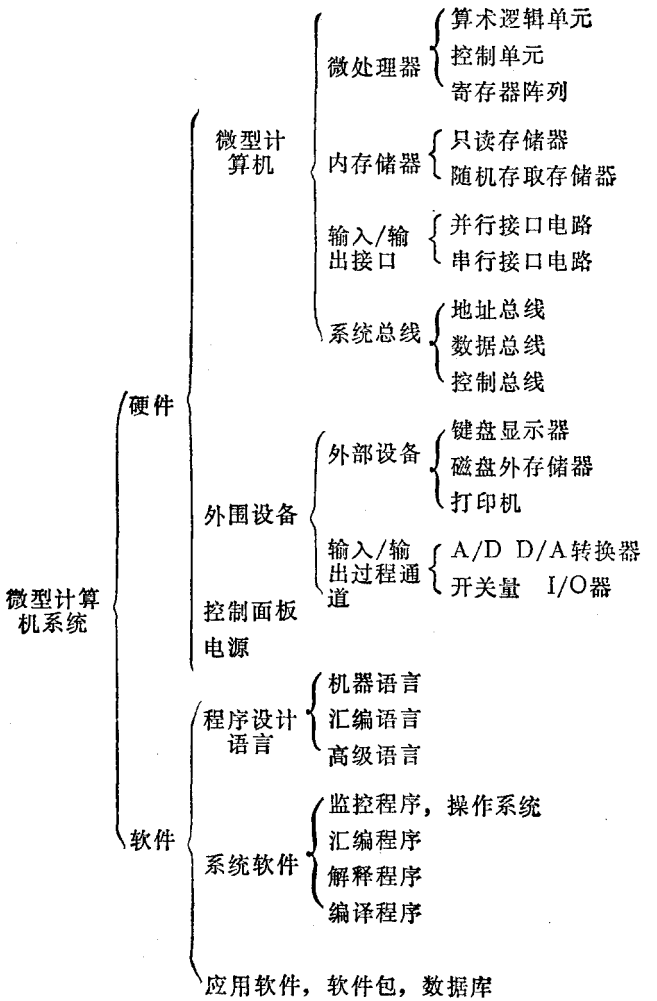
为了能使读者对汉字终端技术有一个比较全面的概括了解，本章从简要介绍微型计算机系统的组成入手，在硬件和软件相结合的基础上，介绍汉字终端设计中应用较多的汇编语言程序设计初步知识，最后介绍汉字终端的几种基本应用软件流程框图。

6·1 微型计算机系统的组成

1. 概述

众所周知，一个完整的微型计算机系统应包括硬件和软件两大部分。所谓“硬件”是指构成微型计算机系统的一切部件和装置的总和，这是计算机实现各种功能的物质基础；所谓“软件”是指为了充分发挥计算机的效能和方便使用者而编成的各种程序和文件的总和，又称为计算机的程序系统。计算机的一切功能都是在软件的控制下实现的，这是计算机系统区别于其它布线组合逻辑系统的一个最显著的特点。

微型计算机系统的组成大致包括以下内容：



在微型计算机系统中，软件和硬件是密切相关的。一方面，进行微型计算机的程序设计，必须与微处理器的结构特点、内存储器地址分配、I/O接口电路配置等硬件相互配合，

通盘考虑。另一方面，开发软件硬化的微型计算机系统，已成为一个重要的发展方向。所谓“软件硬化”，就是把计算机的软件功能固化于硬件中，这种带软件的硬件称为“固件”。把经过调试好了的各种程序固化（写入）于只读存储器(如 MASK ROM和EPROM) 芯片中，就是一种最常见的固件。

2. 微型计算机的硬件

微型计算机的硬件主要有微处理器（CPU）、内存储器（ROM、RAM）、I/O接口电路和系统总线四个部分，它们的组成框图如图6·1所示。

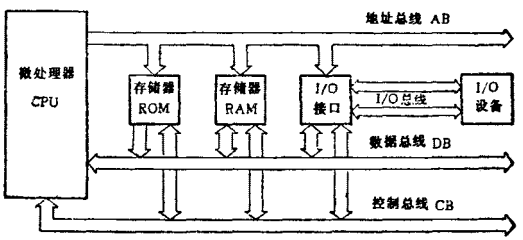


图 6·1 微型计算机系统硬件组成框图

（1）微处理器

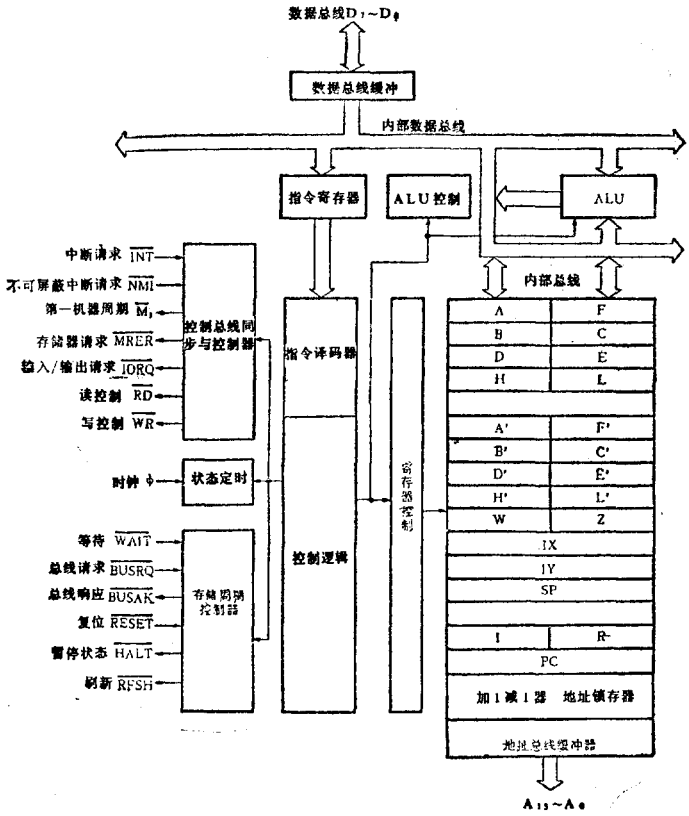
微处理器（μP）又称中央处理器（CPU），是由寄存器、运算器、控制器等部分组成的一块大规模集成电路芯片。微处理器能完成取指令、执行指令以及与外存储器和逻辑部件交换信息等操作，是微型计算机进行运算和控制的核心部件。

微处理器的种类很多，内部结构各不相同，以处理的数据宽度来分，有 1 位、4 位、8 位、16位和32位微处理器。一种常用的 8 位微处理器（Z80）的内部结构和外部引线逻辑图如图6·2所示。

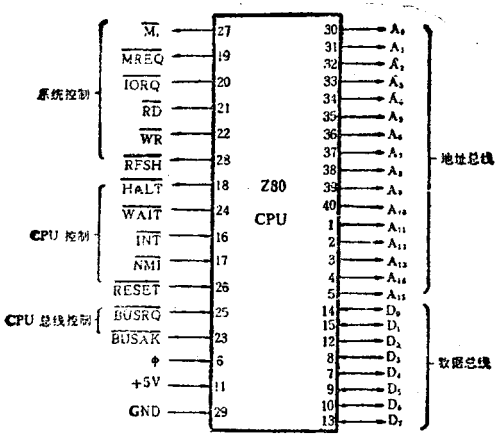
由图（a）可见，微处理器内部大体是由寄存器、运算器、控制器三部分构成的。

① 寄存器

图（a）右下部分，有许多小格的长方形框内是Z 80 CPU的寄存器，包括A～L、A′～L′、W、Z、IX、IY等共18个8



（a）Z 80 CPU 内部结构框图



(b) Z 80 CPU 引线逻辑图

图 6·2 Z80CPU的内部结构和外部引线逻辑图

位寄存器、4 个16位寄存器、2 个 8 位累加器和标志寄存器。

寄存器是微处理器内部的临时存储单元，用来暂时存放运算过程中的地址或数据，也可用以控制程序的执行顺序或存放运算结果的标志。

各种型号CPU的寄存器数量和配置方式各不相同。在汇编语言程序设计中，应根据CPU内部寄存器的配置特点加以利用，借以加快程序运行速度，提高系统效率。

② 运算器

运算器又称为算术运算逻辑部件（图（a）中的ALU）。它的主要功能是对数据进行算术运算和逻辑运算。操作时，运算器通过数据总线取得数据，经过指令规定的运算处理，所得运算结果或留在运算器内以备下次运算时使用，或通过数据总

线写入主存储器。整个运算过程是在控制器控制下自动进行的。

③ 控制器

控制器是微处理器的指挥中心。它按照机器代码程序的要求，控制微型计算机系统的各功能部件协调一致地工作。如程序和数据的输入，内部信号的处理及处理结果输出等操作过程，都是在控制器发出的操作命令指挥下实现的。当CPU进入自动计算状态时，控制器的任务是：从主存储器中取出指令，对该指令进行分析和解释，然后对其它功能部件发出执行该指令所需的各种时序控制信号。接着再从主存储器中取出下一条指令执行。如此连续运行下去，直到程序执行完毕为止。计算机的全部工作过程，就是执行程序中逐条指令的过程。为了实现上述功能，控制器包括图中所示的指令寄存器、指令译码器和定时控制逻辑电路等几部分。此外，还需要一个外接的石英晶体振荡器，用来产生定时信号，控制系统各部分有节奏地工作。

（2）内存储器

内存储器又称为主存储器，是直接连接在系统总线上的半导体存储器（参见图6·1），用来存放微型计算机的程序和数据。根据不同的使用要求，内存储器分为只读存储器（ROM）和随机存取存储器（RAM）两种。ROM主要用来存储微型机的程序和固定的数据；RAM是内存储器的主体，可用来存储临时输入的程序、数据以及在操作过程中的中间运算结果，还可以用作各种数据的缓冲存储器。

内存储器是按地址存取信息的，它通过系统总线与CPU直接交换数据，存取数据的时间与数据所在存储单元的地址无

关。工作时，CPU首先将地址送到存储器的地址寄存器和译码器，同时将“读/写”控制信号送到存储器芯片的读/写控制端。于是，存储器按照地址译码器的输出确定相应的存储地址单元；并按“读/写”控制信号的要求操作：如果是“读”命令，将该地址单元内容送往数据总线；如果是“写”命令，将已存在于数据总线上的数据写入该存储地址单元内，该单元的原有内容自行消失。

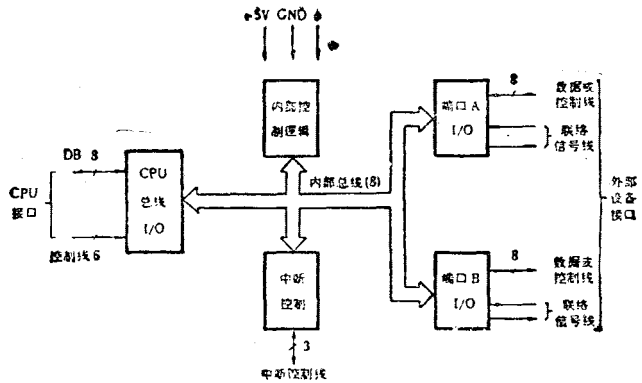
内存储器的特点是存取速度快（与外存储器相比），存储容量大（与寄存器相比），是微型计算机不可缺少的重要部件。内存储器容量的大小，是衡量一个微型机系统性能高低的重要指标。

有关内存储器的内部结构及其它性能特点，已在4·1节的有关内容中介绍，兹不赘述。

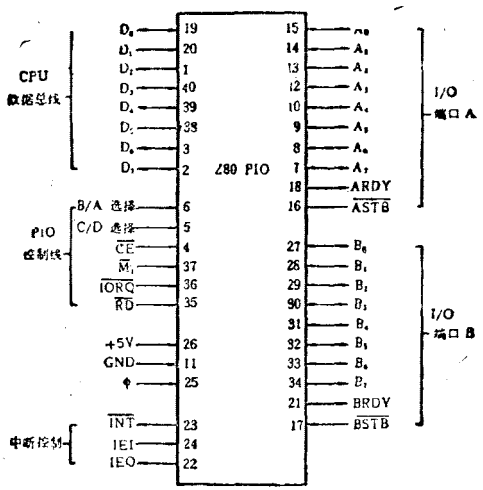
（3）输入/输出接口

输入/输出接口（即I/O接口）是微型计算机与外围设备的连接部件。I/O接口直接连接在系统总线上，CPU通过它向外围设备发送控制命令，控制外围设备的工作过程以及输入输出信息。

外围设备是计算机与外界交换信息的设备，又称为输入/输出设备，简称外设。常用的外设有键盘、显示器、磁盘驱动器、打印机、通信控制器等等。我们知道，CPU是以其本身固有的定时规范和操作特点进行工作的，而种类繁多的外设又都有各自的操作特点，如以人工操作键盘输入信息时的速率可以低达秒级，磁盘存储器的输入信息速率可高达微秒至毫微秒级；又如对高速打印机和显示器的数据传送方式是并行传送，而点阵式打印机和远程通信接口则需串行传送；还有的外设需



(a) 内部结构框图



(b) 外部引线逻辑

图 6·3 Z80 PIO的内部结构及外部接线逻辑图

要进行模拟信号与数字信号的相互转换。这些，都需要通过相应的I/O接口电路，使CPU与各式各样的外围设备协调一致地工作。

由于各种外设的不同特点，I/O接口的种类也很多。如图1·1中的键盘控制器、显示控制器、打印控制器以及通用同步/异步收发器（USART）等等。为了与各种系列的CPU芯片配套，已有许多I/O接口逻辑电路采用了LSI电路芯片，并做到了系列化和标准化。其中有不少I/O接口芯片还具有可编程序功能，能通过程序控制使芯片按照几种不同的方式工作，使用十分灵活方便。现以两种常用的I/O接口芯片为例，说明I/O接口电路的结构及性能特点。

① 可编程序并行I/O接口芯片——Z80 PIO芯片。

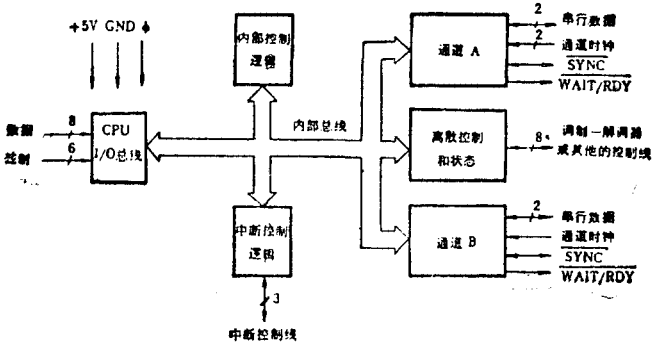
Z80 PIO是为Z80 CPU配套设计的I/O接口芯片。它可与大多数常用的外设直接连接，不必另加逻辑电路；可以进行并行的8位数据传送，或者用于单个控制信号的输入输出（即位控方式）。PIO的内部结构和外部引线逻辑图如图6·3所示。

由图可知，PIO是具有40条引线的LSI芯片。它有两个8位端口与外部设备相接，用来传送数据或控制信号。它还可以根据CPU的程序来设定端口的工作方式。有输出方式、输入方式、双向方式和位控方式四种可以任选。每种方式都配有实现中断所必需的各种信息，可实现中断优先排队，并自动提供中断向量。PIO的所有数据及控制信号线都经系统总线与CPU直接相连，由CPU执行专门的I/O指令来实现数据及控制信号的传送。当与外设之间传送数据时，可以用相应的联络信号线进行数据流量控制，在得到对方允许时才可传送。

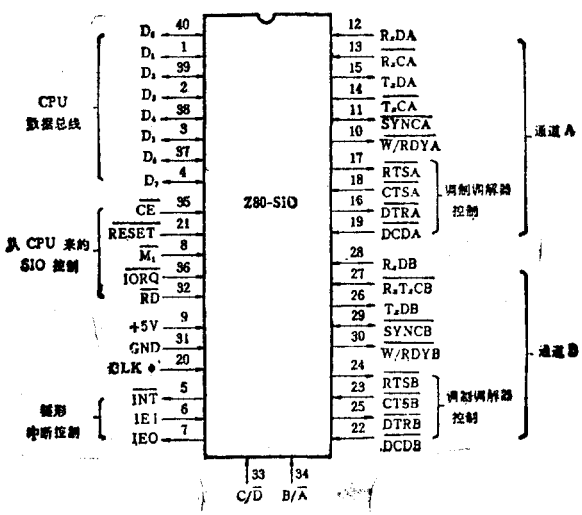
② 可编程串行I/O接口芯片——Z80 SIO

Z80 SIO是为Z80微型机系统进行串行数据通信设计的可

编程双通道接口电路芯片，能适用于异步或同步的通信方式，它的内部结构和外部引线逻辑图如图6·4所示。



(a) 内部结构框图



(b) 外部引线逻辑

图 6·4 Z80 SIO的内部结构及外部引线逻辑图

由图可知，Z80 SIO是具有40条引线的双列直插封装LSI芯片，具有很强的数据通信功能，它能按照数据通信链路的基本型规程或HDLC规程的要求，用程序设定各种工作方式。如同步或异步传输、传输速率、奇偶校验方式等等。SIO的两个通道都有命令寄存器，在工作之前，CPU以一系列命令字来预置所需的工作方式和具体参数，并通过其它一些命令来规定所选工作方式的各种状态，这些过程都由程序的初始化实现的。SIO与CPU的连接方式和中断设置方式，与PIO芯片大致相同。

（4）系统总线

系统总线（BUS）是连接计算机各部件的公共信息通道。CPU、内存储器及I/O接口电路等功能部件，都直接“挂”在总线上，通过总线来传输和交换各种数据信息。为了避免传输数据时互相冲突，凡是挂到总线上的部件（芯片）的数据线，都必须经过三态驱动缓冲电路。平时它们都处于高阻（脱开）状态。需要操作时，由CPU控制，在某一瞬时只允许一个部件打开发送门发送信息。用总线结构方式可以减少传输线的数目，提高系统的可靠性、灵活性。因此，已被各种微型计算机系统普遍采用。

微型机的系统总线分为地址总线、数据总线和控制总线三种，即所谓“三总线结构”（见图6·1），现分述如下：

① 地址总线（AB）

地址总线是CPU用来输出确定的内存储器或I/O接口的地址。它总是单向传输的。一般8位微处理器设有一条16位宽的地址总线，可以指明$2^{16}=65536$个存储单元的地址，也就是说，CPU的寻址范围是64K位。由于LSI电路封装工艺的限制，一

般芯片不超过40～48条引线，为了解决传送能力与引线不足的矛盾，有时还“借”用地址线来传送数据。这时，为了区分传送的是地址还是数据，需用专门的脉冲来选通。

② 数据总线（DB）

数据总线是用于CPU、主存储和I/O接口之间进行数据交换的通道，是双向传输的。数据总线的宽度一般与CPU的“位”数相同，对于8位微处理器来说，它有一条8位宽的双向数据总线；并可通过三态控制门，使CPU及其它部件同数据总线连接或“脱开”，实现数据双向传送。

③ 控制总线（CB）

控制总线用来传送微型计算机内部的各种定时信号和控制信号，以保证各功能部件的操作同步和协调。在控制总线中的数据（即控制信号）传送方向有两种：由CPU发出的控制信号和送入CPU的控制信号。前者有：同步脉冲、定时脉冲、读写操作控制以及应答信号；后者有：准备就绪、中断请求以及保持请求信号。

微型计算机系统总线的具体结构，各计算机厂家都有标准。较常用的有S-100总线、STD总线、SBC总线等等。S-100总线是美国部分计算机制造厂确定的一种标准总线，后作为IEEE-696标准草案推荐，得到较广泛的应用。S-100总线的结构是一块有100条引线的插件板，分成两面，每面有50条引线，其中有16条数据线、24条地址线、8条状态线、5条输出控制线、6条输入控制线、8条直接存储器存取（DMA）控制线、8条向量中断线以及电源线、地线等等。采用统一的总线结构，可对微型计算机的功能扩充带来很大的方便。

3. 微型计算机的软件

微型计算机的软件包括程序设计语言、系统软件、应用软件和数据库四大部分。

（1）程序设计语言

程序设计语言是用来编写计算机程序的语言，通常分为机器语言、汇编语言和高级语言三类。

① 机器语言

机器语言是一种能被CPU直接识别和执行的机器代码。它是由一系列二进制数“0”和“1”组成的代码。每一组代码规定CPU执行某种特定的操作，故称为指令。通常一条指令对应机器的一种基本操作。每台计算机能识别和执行的全部指令的集合，称为该机的指令系统。一般来说，指令系统对不同的CPU是不相同的。

机器语言是计算机唯一能识别的语言，但对程序设计者来讲，使用起来就很不方便。如果用这种二进制代码表示的机器语言来编写程序，不但直观性差，繁琐费时，而且极易出错。

② 汇编语言

汇编语言是一种用助记符来表示的程序设计语言。例如Z80微处理器的汇编语言，用助记符LD（Load的缩写）表示数的传送和装入，用ADD表示相加，用IN和OUT表示对外设的数据输入和输出等等。

汇编语言也是一种面向机器的语言。与机器语言相比，它比较直观，而且容易记忆和检查。当前，在汉字终端的程序设计中，主要还是用汇编语言来编写源程序。但是，CPU不能直接识别和执行用汇编语言写的源程序。需要经过汇编程序加工

（汇编），翻译成机器语言表示的目标程序，才能由CPU识别和执行。

由于汇编语言的语句与机器指令是一一对应的，针对同一问题所编写的汇编语言源程序，对于不同的CPU是互不通用的。在程序的编写难易方面，汇编语言虽然比机器语言大有进步，但仍然比较费时。

③ 高级语言

高级语言是一种面向过程的程序设计语言。上述的机器语言和汇编语言都是与特定的机器有关，称为初级语言。为了从根本上克服初级语言的弱点，使程序设计语言适合于描述各种算法，而不依赖于具体计算机的结构和指令系统，以便于计算机的普及应用，于是就有各种高级语言产生。

高级语言的种类很多，有面向科学计算的公式翻译语言FORTRAN，面向商业的通用语言COBOL，小型通用的交互式会话语言BASIC，面向数值计算过程的算法语言ALGOL，结构程序设计语言PASCAL等等，都是国内外广泛使用的高级语言。近年来在国外流行的LISP、FORTH和C语言引起了人们的重视。LISP是一种具有自编译能力的函数构造式语言，适用于表处理和人工智能。FORTH既是一种高级语言，又是一种汇编语言，适合于过程控制和企业管理。C语言是一种面向结构的程序设计语言，在时间、空间和效率上都能与汇编语言媲美，适用于描述操作系统、编译程序等大型软件，已逐渐成为微型机的主流语言。

用高级语言编写程序，人们可以不必化费精力去了解计算机的内部逻辑，而把精力集中于算法和过程的描述；而且由于这种语言比较接近人们的习惯，易学易用易维护，因此，使用高级语言能大大提高编程效率，有利于促进计算机的普及应用。但

是高级语言必须经由编译程序或解释程序翻译成目标程序，机器方能执行。与汇编程序相比，编译程序和解释程序所需的内存容量比较大。

（2）系统软件

系统软件是为了充分发挥计算机效能和方便用户而提供的一系列程序和文件的总称。它主要包括监控程序、操作系统、汇编程序、解释程序和编译程序等。

监控程序又称管理程序，是为合理使用计算机资源和方便用户而设计的一套简单的系统软件。其主要功能有：合理安排主机及外部设备的操作；按轻重缓急处理各类中断；接受和分析各种命令，实现人机联系；控制源程序的编译、装配、启动等。监控程序通常直接固化在EPROM中，用来管理单板计算机的运行。

操作系统是在监控程序的基础上进一步扩充而成的大型程序系统。其主要功能有：组织整个计算机的工作流程；管理和调度各种软、硬件资源；检查程序和机器的故障；实现计算机资源的多用户共享等。所谓“计算机资源”，通常是指存储器、处理器、外围设备和信息四大类。这些资源都需用操作系统进行管理和调度。一个用户程序只有通过操作系统获得所需的资源以后，才能被计算机执行。操作系统还向用户提供控制机器的各种命令，并且执行用户用键盘或其它方式输入的命令。因此，计算机配置操作系统后不但可大大提高机器效率，而且使用户的操作使用更为方便。微型计算机操作系统的类型和版本很多，通常存放在磁盘中，开机时调入内存储器，容量一般为十几K到几十K字节。

汇编程序是把汇编语言写的源程序翻译成目标程序的一种

软件。汇编程序有自汇编和交叉汇编两类。前者需要有足够大的存储容量来存放自汇编程序；后者可借助于存储容量较大的一台主计算机，由主机的汇编程序翻译成目标程序，然后送入本机执行。

编译程序是把高级语言编写的源程序翻译成机器语言的目标程序、或翻译成汇编语言的一种翻译程序。它的实现算法较为复杂，常用于翻译规模较大、复杂性较高、而且需要高效率运行的高级语言书写的源程序。

解释程序是按照源程序的指令或语句的动态执行顺序，逐条或逐句翻译并立即执行的一种翻译程序。它的特点是把源程序直接翻译成输入数据，并产生所要求的动作，而不是产生机器语言形式的目标程序。

解释程序的突出优点是实现算法简单，而且易于在解释执行过程中灵活方便地插入修改和调试措施。但解释程序的最大缺点是执行效率很低，适用于调试用的系统及交互式会话语言（如BASIC）的翻译和执行。

（3）应用软件

应用软件是专门为解决某个应用领域里的具体任务而编制的程序。随着计算机应用的日益广泛，人们已经编制出大量的各种应用软件，如生产过程控制、数据处理、联机通信等等，可按其功能组成适用于不同机型的各类软件包，减少大量重复劳动。我们将在下节介绍的汉字终端输入输出程序、字库管理程序、文本编辑程序等，都是属于应用软件的范围。

（4）数据库

在计算机系统中需要处理大量的数据，检索和建立各式各

样的表格。将这些数据和表格按一定的形式和规律加以组织，建立数据模型，实行集中管理，这就是建立数据库，对数据库进行组织和管理的软件称为数据库管理系统。八十年代以来，数据库技术发展十分迅速，并且已在信息处理、情报检索和各种管理系统中得到广泛的应用。

以上概略介绍了计算机各种软件的内容和特点。丰富的软件是对计算机硬件功能的有力扩充，使计算机系统的性能更强、可靠性更高、使用更方便。在汉字终端设计中涉及的软件知识，主要是汇编语言程序设计和有关的应用软件设计方法。

6·2 汇编语言及程序设计初步

1. 概述

我们要进行汉字终端的程序设计，就要应用微型计算机的程序设计语言。如前所述，程序设计语言有机器语言、汇编语言和高级语言，究竟采用哪种语言合适？这要取决于微型计算机的性能、特点和应用场合。

机器语言是计算机唯一能识别的程序设计语言，也就是机器的指令系统。但它是由一连串冗长的二进制数“0”和“1”表示的，人们很难直接用它来编写程序。

高级语言是面向过程的程序设计语言。它的语法语义与人们的习惯语言相近，用来编写程序比较直观灵活方便，可以大大提高编程的工作效率。而且，由于用高级语言编写的程序，与计算机的具体结构无关，通用性强。可以避免大量的编程重复劳动。因此，各种高级语言的出现，对于计算机应用的普及

和推广，有很大的促进作用。但是，用高级语言编程占用计算机的内存容量大，执行时间长。这是因为，高级语言编写的源程序，需经过编译程序翻译成机器语言的目标程序才能被执行。而编译程序本身需要占用16～32K内存容量；而且用编译程序编译出来的目标程序要比“汇编”出来的目标程序长1～2倍。这不但加大了内存容量，还使目标程序执行时间延长约1～3倍。

汇编语言是把指令、地址和数据都加以符号化的初级语言，因此它比机器语言易懂、易记忆和易于修改。在用汇编语言编程时，可以直接操作到CPU内部的寄存器，把运算过程安排得非常具体，因而能编制出十分紧凑而有效的程序来。能做到既节省内存容量，又提高机器执行速度。也就是说，用汇编语言编程可以做到精细地控制和使用机器资源，在空间和时间上充分发挥计算机的潜力，因此常用于系统程序、实时控制程序和标准子程序的设计。在专用的微型机系统如汉字终端、智能仪表的设计中，汇编语言也是常用的程序设计语言。

但是，用汇编语言编写程序比用高级语言费时费力，因为汇编语言是与机器的具体结构和指令系统密切相关的，程序员用汇编语言编程时，必须具备一定的硬件知识，要熟悉机器的内部结构、指令系统及寻址方式等等，方能进行程序设计。因此而带来的另一个缺点是通用性差，就是在这一机器上设计好的汇编语言程序，不能简单地移植到另一机器上使用。

2. 指令结构及其功能

如前所述，计算机的全部操作都是在程序控制下进行的。程序就是为解决某一具体问题而编写的一系列指令。它通过每一条指令来控制计算机的每一步具体操作。要用汇编语言编

程，就必须掌握该计算机的指令结构及其功能。

指令是指示CPU执行某一特定操作的命令。它是由二进制代码形式表示的，其基本格式如图6.5所示。

图 6·5　指令的基本格式

如图所示，一条指令的基本格式包括操作码和地址码两部分。而指令的具体格式则取决于这两部分的不同内容，现分别说明如下：

（1）操作码部分

操作码是用来指定CPU执行某一种具体操作，8位微处理器一般有几百种操作功能，可大致分为以下四类：

① 数据传送操作

数据传送操作是在CPU内部寄存器之间，或CPU与存储器、I/O接口之间传送数据。即将源单元的内容复制到目标单元中去，而源单元的内容不变。

② 数据处理操作

数据处理操作是把CPU内部寄存器或存储器中的内容进行运算处理。操作后上述单元内容将发生变化，而运算结果的特征将保留在标志寄存器中。运算操作的功能包括：算术运算、逻辑运算、移位、比较、“取反”、“取补”等内容。

③ 程序控制

程序控制是指改变CPU执行指令顺序的操作。在一般情况下，CPU是用程序计数器(PC) 加1的方法来逐条顺序执行指令的。程序控制操作指令可以改变PC的正常值，从而使程序

转到重新指定的起始点执行。程序控制操作包括条件转移和无条件转移、调用和返回、中断控制等内容。

④ CPU控制

CPU控制是用来控制CPU本身的工作状态，如暂停、中断、允许或禁止、空操作等。

（2）地址码部分

地址码是用来指明操作数存放单元的地址。当CPU执行数据传送或数据运算操作时，被传送或参予运算的数据称为操作数。有的指令需要一个操作数（如“传送”、“取补”等），有的指令需要两个操作数（如进行算术运算或逻辑运算时）。用指令的地址码部分来指明操作数存放地址的方式，称为寻址方式。各种型号的微处理器寻址方式不尽相同，寻址方式的多少是衡量微处理器功能强弱的一个重要指标。Z 80微处理器的寻址方式比较多，大致可分为以下三类：

① 立即寻址方式

立即寻址方式就是把实际的操作数（称为立即数）直接写在指令的操作码之后（即地址码部分）。CPU就用这个立即数进行操作，不再需要去寻找它的存储地址了。

立即寻址方式的立即数为一个字节（8位），如果立即数是两个字节，就称为立即扩展寻址方式。

② 寄存器寻址方式

寄存器寻址方式就是以CPU内部的某一寄存器内容为操作数。由于CPU内部寄存器数量较少，只需较少位数的地址码便可指明。例如有三位地址码便可指明8个内部寄存器。因此寄存器寻址方式的指令比较短（单字节），再加之内部寄存器存取操作速度快，故寄存器寻址指令的执行速度比较快。

寄存器寻址方式有两种扩展的形式：一种是隐含寻址，就是在指令中没有直接指明是哪个寄存器的内容参加操作，而是靠操作码来隐含表示。通常是用寄存器A（即累加器）作为隐含值。

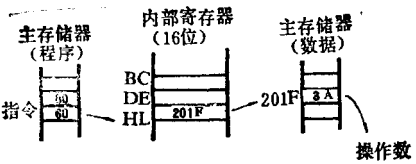


图 6·6　寄存器间接寻址示意图

另一种是寄存器间接寻址，就是以寄存器的内容作为操作数存储单元的地址，如图6·6所示。设有操作数“3A”存储在“201F”地址单元内，而地址数“201F”又存放在内部寄存器HL内。在间接寻址方式的指令中只需指明内部寄存器HL，便可找到操作数“3A”。

由于HL寄存器是双字节长的16位寄存器，故可表示的地址数有 $2^{16}=65536$ 个，换句话说，可寻址范围为64K。

③　直接寻址方式

直接寻址就是在指令的操作码部分之后，直接指明了16位地址码。这个地址码可以是操作数的存储地址，也可以是程序的转移地址。具体属于哪一种，则由操作码指明。

直接寻址方式也有两种扩展形式：一种叫变址寻址，就是地址码部分给出的是变址寄存器中基地址的位移量；另一种叫相对寻址，就是地址码部分给出的是相对于现行程序计数器（PC）地址的位移量。

上面提到的变址寄存器是CPU内部的16位寄存器，在图6·2(a)中通常是指寄存器IX、IY。程序员预置在IX、IY寄存器中的16位数，称为变址寄存器的基地址。

变址寻址和相对寻址方式的示意图如图6·7所示。

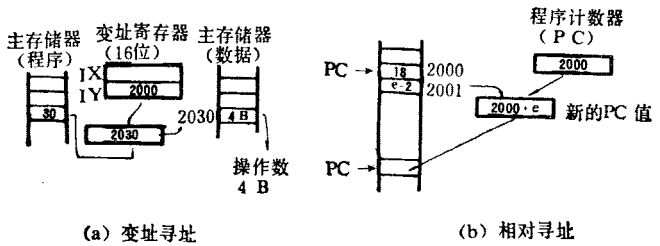


图 6·7 直接寻址的扩展形式示意图

图(a)为变址寻址方式，操作数4 B存放在主存储器2030单元内。在变址寄存器IY内已经存放着基地址数2000，当程序执行到需要读出操作数4 B时，在指令的地址码段只需给出地址位移量30，便可在2030单元读出操作数。图(b)是相对寻址方式，当程序计数器PC为2000时，执行到一条指令18 e-2（双字节指令），前一个字节18是操作码，说明这是一条相对寻址指令；后一字节中的e是地址位移量，说明要把PC指针转移到现行值加位移量（2000 + e）的新地址。

应用直接寻址的扩展形式来寻址，可以使指令的地址码段缩短为一个字节，节省了程序存储器的内存空间，而且便于程序的调试和运行。

3. 汇编语言

如前所述，汇编语言是一种用符号表示的程序设计语言。它把每条指令中的操作码、地址码或数据都用字母数字符号

（称为助记符）来代表，以便于编程和调试。但是，用汇编语言写的源程序，必须经过汇编程序翻译成机器语言的目标程序，才能被机器所接受，它们三者的关系如图6·8所示。

图 6·8　汇编语言、汇编程序与机器语言的关系

汇编语言是面向机器的语言，不同型号的微处理器所用的汇编语言是不相同的，但在语法规则方面有很多相似之处。现以Z80微处理器的汇编语言为例，来说明汇编语言的一般规则。

（1）汇编语句的构成

汇编语言的源程序是由语句组成的。这些语句称为汇编语句或汇编指令，一般可以分成四个部分（或四个字段），其格式如图6·9所示。

标号	操作码	操作数（地址）	注解

图 6·9　汇编语句的基本格式

现分别介绍每个字段的含义和规则。

①　标号字段

标号通常由程序员自定的字母、数字及一些其它符号组成。但标号的第一个字符必须是字母，其长度一般限制在5～6个字符以内。

标号的作用是给指令的地址或数据起一个符号名字，以便在其它语句的操作数字段中引用。标号是任选的，一般语句可

以不要标号，只有当这条指令语句成为其它指令的转向目标时，才有必要给这条指令规定标号。一个特定的标号在程序中只能被定义一次，否则便会在转向时出现二义性的错误。标号一般以冒号“：”结束，有的机器则规定可以不用冒号。

② 操作码字段

操作码是语句的基本字段，它是用助记符表示的机器操作指令。汇编语句中其它字段都可以省略，但必须有操作码字段。

汇编语句中的操作码助记符一般用接近人们习惯语言中的字符来表示，但不同型号的微处理器规定又各不相同，如“数据传送”操作，在Intel 8080A中的助记符是MOV（move的缩写），在Z80中的助记符是LD（load的缩写）。因此使用汇编语言编写程序时，必须先认真熟悉所用CPU的编程手册。

有些汇编语句中的操作码字段并不表示CPU的操作指令，而是用于向汇编程序说有关信息的“伪指令”，我们将在下一小节中进一步说明。

③ 操作数字段

操作数字段又称为地址字段，是用来指明操作对象的名称或数据。它可能是寄存器号或存储器地址，也可能是一个符号或标号；可能是一个算术表达式或逻辑表达式，也可能是一组十六进制的、十进制的或二进制的数据，具体用哪一种，完全由指令的操作要求决定。例如，在Z80汇编语句中，LDA,B就是表示把寄存器B中的数据传送到寄存器A（即累加器）中。如果操作数是十六进制或二进制的数据，则应分别注明“H”或“B”，例如2DH，01100110B等。若是字母开头的十六进制数，还应在字母的前头添“0”，例如0A1H，

Retro Workshop 果粉工作室

0 FFH等等。

④ 注解字段

注解是用来对程序进行说明的一段文字，以便于人们阅读。它必须以分号“；”开始。如果一行写不下需要另一起行时，也必须以“；”开始。汇编程序对于“；”以后的内容一概不予处理，当然就不会出现在目标程序中了。

语句有了注解字段，可以使程序更清楚，便于阅读和调试。但是，并不是每条语句都需要进行注解。

（2）常用的伪指令

汇编语句中有些操作码字段，被用于向汇编程序提供必要的控制信息，以执行一些特殊的操作。例如：为程序指定一个存储区；把一些表格和数据放入存储区；定义一些符号；表示程序的结束等。由于这些操作码只是告诉汇编程序该做些什么，并不象机器操作指令那样需要翻译成机器代码由CPU执行。因此在汇编以后的目标程序中已不复存在。人们就称这些特殊的操作码为“伪指令”，或“汇编命令”。

汇编语言的伪指令较多，现举若干常用的伪指令为例加以说明。

① 起始地址说明指令ORG，用于向汇编程序说明目标程序的起始地址。如有语句：

```
ORG  2000H
```

说明以下一段目标程序的起始地址为2000H单元。

② 符号赋值指令EQU，用于向汇编程序说明标号的具体数值。如有语句：

```
CR:  EQU 12
```

说明标号CR的值等于12。在这条语句以后的源程序中凡是出现

CR字符时，汇编程序都把它看作是十进制数12。

③ 定义数据字节指令DB，用于确定在程序中存放到存储区的数据。该数据可以是单字节的，也可以是一个字符串。例如有语句：

STR：DB 54 79 3D

就是告诉汇编程序把数据54 79 3D存放在以标号STR为起始地址的三个存储单元中。

④ 程序结束指令END，这是通知汇编程序该源程序结束的伪指令，因此必须安排在程序的最末一条。有时在一段程序的末尾加上一条HALT语句，这是告诉CPU进入“暂停”状态的管理指令，它属于机器操作指令，不是伪指令。

（3）汇编语句的应用

我们用一个简单的计算程序实例，来说明汇编语句的具体应用。

【例】 求两个8位数（3AH，2BH）之和，将结果存放在2100H单元内。

用Z80汇编语言编写的源程序为：

序号	标号	操作码	地址码	注　　解
1		ORG	2000H	；程序从2000H单元开始
2	RSL：	EQU	2100H	；定义结果存放单元标号
3	ST：	LD	HL，ADR	；数据地址单元装入HL
4		LD	A，(HL)	；数据送入累加器A
5		INC	HL	；地址单元数加1
6		ADD	A，(HL)	；两个数据相加
7		LD	(RSL)，A	；结果送入规定单元
8	ADR：	DEFB	3A2BH	；定义数据存放区
9		END		；结束

这是一段有9条汇编语句的源程序，其中第1、2条和第

8、9条是伪指令，用来控制机器汇编程序的；第3～7条是机器操作指令。

① 伪指令的应用

第1条（ORG 2000H）是告诉汇编程序，目标程序在内存中存放的首地址是2000H单元。也就是确定了程序开始执行时的地址单元。

第2条（RSL:EQU 2100H）是给计算结果存放单元2100H定义一个标号RSL。凡在以后的语句中出现字符串RSL，汇编程序即认为是2100H单元。

第8条（ADR: DEFB 3A2BH）是给两个8位数确定存储单元，也就是将这两个数存放在ADR和ADR+1的存储单元内。ADR的地址值是多少？可以从机器操作指令（第3～7条）所占的内存单元算出。设上述指令占9个内存单元，则ADR的单元地址为:2000H+9+1=200AH单元。

第9条（END）是程序结束指令，表示本段程序的终止。

② 机器指令的应用

第3条（ST：LD HL，ADR）是直接寻址方式的数据传送指令。它通知CPU执行下述操作：把ADR地址单元（即200AH）送入16位寄存器HL。这条语句有一个标号ST，是为了便于在以后的程序中引用，当然，在这个简单程序中，这个标号也可以省略不用。

第4条（LD A，(HL)）是寄存器间接寻址方式的数据传送指令。它通知CPU把HL寄存器中的地址单元存放的一个8位数（即3A）送到累加器A中。在编写这类指令时，必须注意把存储单元的地址（如200AH）与该地址单元内存放的内容（如3A）区别开。

第5条（INC HL）是16位数加1的算术运算指令。它通知CPU把HL寄存器的内容加1，即指向200BH地址单元。

第6条（ADD A，(HL)）是寄存器间接寻址的加法指令。它通知CPU把寄存器HL中的地址单元存放的一个8位数（即2BH）与累加器A中的8位数（即3AH）相加，和数放在A中。

必须注意，由于执行了第5条指令，这时的（HL）所指的内容与第4条中（HL)所指的内容已不同了：一个是2BH，一个是3AH。

第7条（LD（RSL），A)是直接寻址的数据传送指令。它通知CPU将累加器A中的内容（两数之和）送到指定的地址单元RSL（即2100H）中去。

以上介绍了汇编语言的初步知识，作为学习汇编语言程序设计的入门响导，读者还需通过在实践过程中从易到难、由浅入深的不断练习，才能逐步做到真正掌握和应用。

4. 程序设计初步

要应用微型计算机完成一项确定的任务，必须首先安排实现该任务的详细计划和具体步骤，并运用机器能够识别的语言加以描述。这就是微型计算机的程序设计。为了进行程序设计，就要对任务的性质和特点进行详尽透彻的分析，然后建立相应的数学模型，选择适当的计算方法。就汉字终端的软件技术来说，主要涉及应用CPU实现汉字和字符代码的转换及处理，进行各种数据代码的逻辑运算和算术运算，基本上不涉及复杂的数学问题。所以，程序设计的主要内容就是在深入分析各工作单元特性的基础上编制流程框图，选用合适的程序设计语言，然后按照流程图每一框图内的要求，编写出源程序。如

果是用汇编语言编写的源程序，则经过汇编程序（或人工汇编）翻译成机器能识别和执行的目标程序，在机器上进行调试和运行，最后经修改正常后才告完成。

程序设计过程的示意图如图6·10所示。

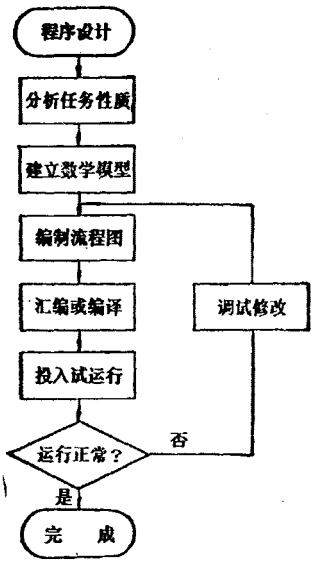


图 6·10　程序设计过程示意图

在上述的程序设计过程中，编制流程框图是其中关键的一环。

编制流程框图就是以约定的几何图形及必要的文字说明来描述微型计算机的计算过程。通过流程框图可以使人们对程序的基本过程一目了然，并有利于将较大的程序分成模块，进行分工设计，最后综合调试运行。

如前所述，程序是若干条指令的有序集合。程序的基本执行过程是顺序地逐条执行指令。但是，在实际的程序设计中，

除了最简单的程序外，始终保持顺序执行的程序是极少见的。本节主要应用流程框图的方法来着重介绍分支程序、循环程序、子程序的基本概念。

（1）分支程序

在程序设计中，会常常遇到各种条件判断和比较操作，如：相等或不相等、大于或小于、结果为0或结果为1等等。CPU根据这些判断或比较的结果（即“条件”）来决定程序的执行顺序，就形成了分支程序。分支程序是常用的程序结构形式之一，它体现了计算机的智能特点。

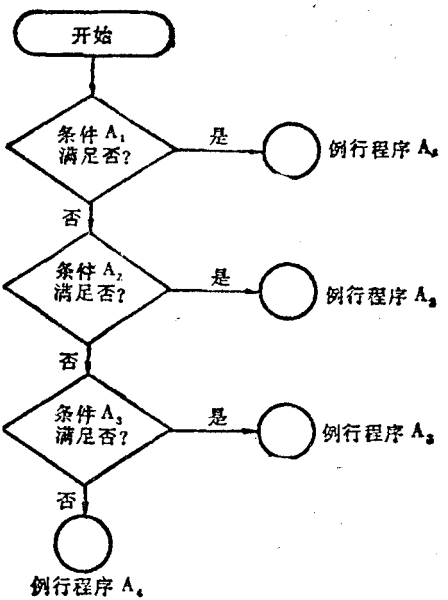


图 6·11 简单的分支程序框图

进行分支程序设计的关键，是要正确掌握标志寄存器（参见图6·2(a)中的“F”）的标志位变化，使程序按照设计要求实现正确转移。

图6·11是一个具有三次条件判断操作的简单分支程序框图。

如图所示，当程序开始后的第一次判断条件满足时，就转移到执行例行程序A_1；否则就顺序往下执行，进行第二次判断。第二次判断满足条件时，就转移到执行例行程序A_2；否则就顺序执行进行第三次判断……。如果全部条件都不满足，就执行例行程序A_4。

在实际工作中，如果遇到多次判断或比较的情况时，实现程序转移的更有效的方法是使用转移表。所谓“转移表”就是把各个例行程序的首地址以列表的形式存放在一个预定的存储区中，一个首地址占用两个字节，其示意图如图6·12所示。

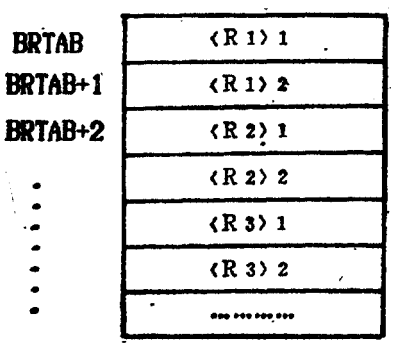


图 6·12　存储区中的转移表示意图

图中，左侧标出的是转移表的地址号。该行第一个标号“BRTAB”是转移表的首地址，具体地址数由程序决定。图中方框里的标号是各例行程序的首地址，每个例行程序占2个

字节。如<R 1> 1、<R1> 2，<R2> 1、<R2> 2，…… 等。当主程序执行中满足判断条件需转移到第n个例行程序时，只要把程序转向BRTAB + 2(n - 1)的地址单元执行就行了。

（2）循环程序

在实际应用中，常遇到需要多次反复执行某些相同操作的情况，这时，应用循环程序可以使程序设计过程简化，并且可以节省内存。一般循环程序框图如图6·13所示。

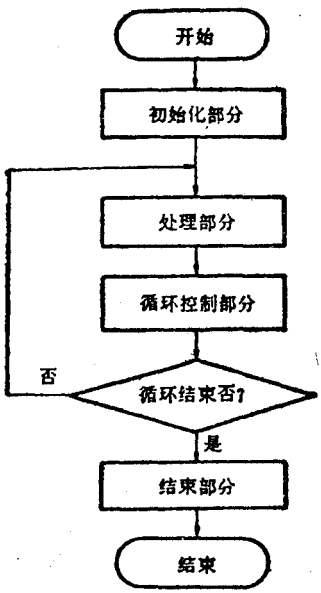


图 6·13 一般循环程序框图

由图可知，一般循环程序由以下四个部分组成：

·初始化部分，用来设置循环初态。它包括循环工作部分初态（即变量初值）和循环结束条件初态（即预置计数器的循

环次数）。

·循环处理部分，即需要反复执行的程序段，这是循环程序的主体部分。

·循环控制部分，用来实现对循环的判断和控制。它包括修改计数器的计数值以及对循环结束条件的判断，这是循环程序设计的关键部分。

·循环结束部分，即存储或输出运算结果。

（3）子程序

在程序设计中，把具有相对独立性的或需多次反复应用的程序段编成子程序，这是提高编程效率、节省内存空间的重要手段。

任何一个具有入口地址和以RET（返回指令）结尾的程序段都可称为子程序。调用这个子程序的程序称为主程序，它通过CALLnn（调用指令）来调用子程序。调用过程示意图如图6·14所示。

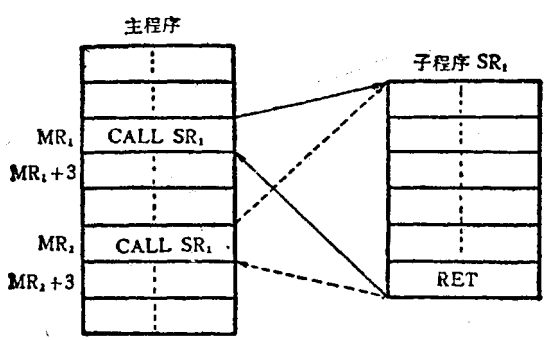


图 6·14 主程序调用子程序示意图

图6·14中，当主程序执行到MR_1地址单元时，程序中出现

CALL SR_1(调用子程序SR_1)指令。这时程序计数器PC立即指向被调用子程序的入口地址SR_1，并从该地址单元开始顺序执行子程序。一直到最后执行RET（返回指令）时，PC又自动返回到调用指令的下一个地址单元MR_1+3开始执行（因为CALL SR_1是三字节指令）。到MR_2时，又进行第二次调用。

由此可见，子程序必定以RET指令结尾，使子程序执行完后能自动返回主程序继续往下执行，这是子程序与分支程序中例行程序的主要区别。

6·3 汉字终端的应用软件

如前所述，汉字终端是在已有的微型计算机软硬件基础上设计而成的，因此，进行汉字终端的应用软件设计时，应深入分析掌握原型机的软硬件资源，要熟悉原型微机(包括单板机)的操作系统或监控程序中的输入输出程序、文件管理程序等等，并加以充分利用，才能使汉字终端的应用软件设计收到事半功倍的效果。

汉字终端的应用软件是面向具体应用而编制的一系列程序的总和。由于各种汉字终端的使用要求不同，应用软件的种类很多，不可能一一列举。这里主要讨论几种通用的应用软件，如键盘输入管理、汉字打印和显示输出、汉字文本编辑等程序的基本流程框图以及初步的设计方法。

1. 键盘输入管理程序

键盘输入管理程序的主要作用，是把操作者击键时所给定的键盘码值（座标值）转换成机器的相应内部码。

如前所述，当使用标准键盘输入西文字符时，CPU通过

对键盘按键开关的检测并调用键盘输入管理程序，就可以完成一个字符的输入。当使用标准键盘进行汉字输入时，通常需要从键盘上输入多个字符才能表示一个汉字。也就是说，CPU在调用键盘输入程序完成一个字符的输入之后，还要检测是否已经把表示一个汉字的全部字符输入完毕。如果没有输入完，就返回输入状态继续等待输入；如果输入完毕，要检测一下输入代码是否有错。如果有错则又返回输入状态等待重新输入，并给出相应指示；如果没错就调用输入代码转换程序，进行汉字输入码到内部码的转换，完成一个汉字的输入过程。由此可见，标准键盘的输入汉字管理程序比输入西文字符的管理程序复杂。

（1）标准键盘汉字输入程序的基本流程

标准键盘汉字输入程序的基本流程图如图6·15所示。现逐项说明如下：

输入程序初始化的主要内容是置输入数据缓冲区的首地址以及使输入设备处于准备好状态。

接着是确定汉字输入编码方案。一般汉字终端配有多种输入编码方案供使用者选择（图中画出了六种输入方案）。除国标码和区位码外，常用的还有电报码、汉语拼音、双拼编码以及一、二种字形编码。在按照西文字符相同的输入方法读入一组字符数据以后，程序对读入的字符进行正确性检查。

输入检查的内容是按不同输入编码方案的规则来核对码长及输入字符是否有效，如果有错转入出错处理程序。输入检查正确后就进行代码转换，把输入码转换成机器内部码。若有重码字，则将全部重码字显示在显示器的提示行处，供操作者进行二次选择（再输入一个数字）。若没有重码，则把汉字显示

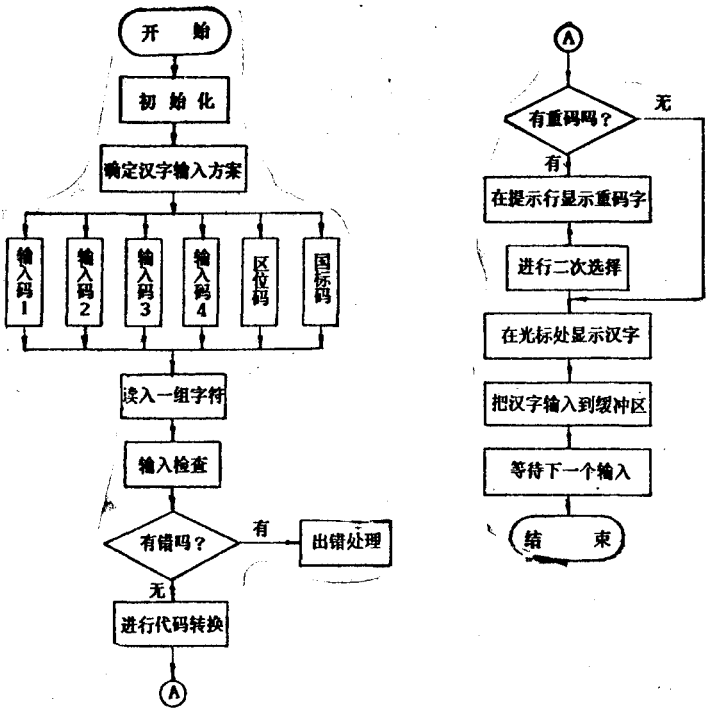


图 6·15 汉字键盘输入程序的基本流程图

在显示器的光标处，并把该字的内部码送入指定的输入缓冲区存放起来，并等待下一个汉字的输入。

（2）输入代码转换

进行输入代码转换就是把不同输入编码方案的汉字输入代码转换成相应的机器内部码，这是汉字键盘输入程序的关键。

如前所述，有些汉字系统的输入编码与内部码是一致的

（如1011汉字自动打字机的见字识码方案、APPLE机汉字系统的仓颉天龙方案等）当然就不需进行输入码到内部码的转换了。

用国标码或区位码方案输入时，从输入码到内部码的转换比较简单。因为绝大多数汉字系统的内部码设计，与国标码有着简单的对应关系，如大多数机器采用的双字节内部码方案，是在国标码的两个字节的最高位置“1”，则当用国标码输入时，只需将接收下来的两个字节的最高位加“1”，就得到相应的内部码了。例如：汉字“大”的国标码为34—73两个字节，在最高位加“1”后，就转换成B4-F3两个字节的内部码了。如图6·16所示。

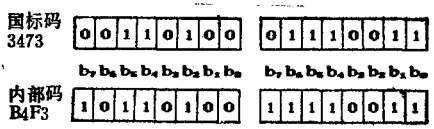


图 6·16 “大”字的国标码和内部码

用区位码输入时，可先将区位码用简单算法变换成国标码，再按国标码输入的方法处理。例如，汉字“大”的区位码是20—83，先把它转换成十六进制数是14—53，于是：

“大”的国标码为 $(1453)_{16}+(2020)_{16}=(3473)_{16}$

内部码为 $(3473)_{16}+(8080)_{16}=(B4F3)_{16}$

由此可见，用区位码输入时，从输入码到内部码的转换，只需用简单的算法便可实现。

除了上述两种输入方案以外，其它的汉字输入编码到内部码的转换，通常不能用简单的算法来实现，而是需要采用查对照表的方法。因此，必须按照不同的输入编码方案编制出输入码与内部码之间的对照表，并且选择适当的查表方法。

要正确编制汉字代码对照表，首先要确定汉字输入码与内部码之间的结构关系。然后选择一种比较适合于这种结构关系的代码对照表。

对照表的具体编制方法很多，大体上可分为三种类型：第一种是以输入编码的代码递增或递减顺序排列的对照表，如表6·1所示。

表6·1

汉 字	笔形码	内 部 码
到	16	B5BD
国	25	B9FA
出	26	B3F6
在	31	D4DA
行	33	D0D0
八	34	B0CB
力	35	C1A6
九	36	BEC5
千	37	C7A7

第二种是以内部码的代码递增或递减顺序排列的对照表，如表6·2所示。第三种是以输入编码的顺序排列，另外又加索

表6·2

汉 字	内 部 码	笔形码
八	B0CB	34
出	B3F6	26
到	B5BD	16
国	B9FA	25
九	BEC5	36
力	C1A6	35
千	C7A7	37
行	D0D0	33
在	D4DA	31

引表的方法。这三种对照表的输入编码转换程序流程如图6·17所示。

（3）查表方法

由图6·17汉字输入编码转换流程图可知，不论采用哪种代码对照表，查表程序是不可缺少的一步。根据汉字代码对照表的不同类型，大体有以下两种不同的查表方法。

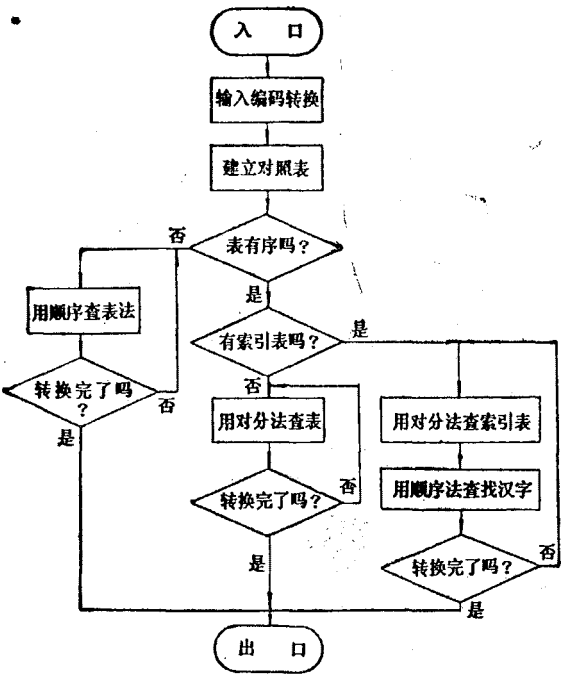


图 6·17 汉字输入编码转换流程图

① 顺序查表法

假设用笔形编码法输入汉字“在”（笔形码为31），需将

它转换成内部码输入指定内存单元。如果对照表是按照内部码为序编制的（如表6·2）则需用顺序查表法，其程序流程图如图6·18所示。

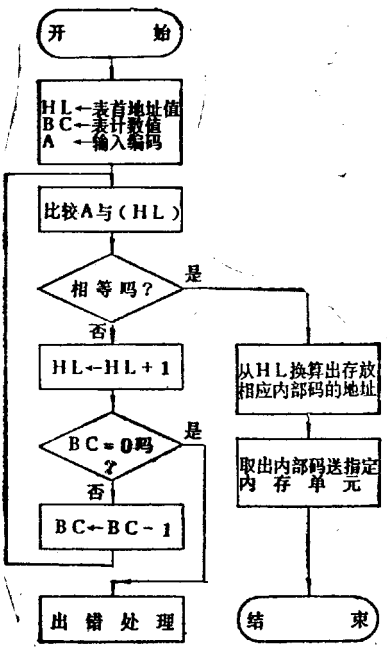


图 6·18　顺序查找法流程图

图中HL、BC是Z80CPU的内部16位寄存器，A是累加器。程序开始后的第一步（第1框），相当于初始化，第2～6框是循环程序的主体，由BC寄存器作为计数器控制循环次数。若一直到BC＝0仍未查到（即A与HL里的内容逐个比较都不相等），则转入出错处理，表示无此输入编码；若查到此数，则把HL的当前值换算成存放相应内部码的地址单元值，然后

取出内部码值送到指定的内存单元，程序结束。

顺序查表法适用于汉字以内部码或国标码为顺序排列的情况，此法的查表操作次数多，速度慢，如在上例中，需要查找比较 9 次才能找到。当对照表项数比较多时，则使输入转换效率大为降低。

② 对分查表法

对分查表法又称折半查表法，是查找有序表的常用方法。用此法查表，可以大大提高输入效率。

假设有一个按照输入字形编码的代码递增顺序排列的汉字代码对照表（参见表6·1），当输入“在”字的笔形码（31）以后，用对分查表法把输入码转换为内部码的程序流程图如图6·19所示。

流程图中BC、DE、HL和A都是Z80CPU的内部寄存器和累加器代号。程序开始后，首先把表6·1在内存中的首地址和末地址分别存放在BC和DE寄存器中，这是程序的初始化。接着把输入的汉字“在”笔形码31放入累加器A中，于是就开始查表操作了。

按照对分查表法的要求，把累加器A中的内容首先与表的中间一项即第 5 项的笔形码33进行比较。如果两数相等，一次查找成功，就可以算出相应的内部码地址单元，读出输入汉字的内部码，结束一个汉字的输入过程。如果两数不等，就需判断要查找的码值是在表的上半部还是下半部：若判断在下半部，即A＞(HL)，则末地址不变，把首地址值换成中项的地址值，再重复一次查找过程；若判断在上半部，即A＜(HL)，则首地址不变，把末地址值换成中项地址值，继续重复查找过程。图6·19中的两个循环程序流程就是按照上述思路设计的。循环程序流程的最后一项DE－BC≤1是循环控制，当被查找

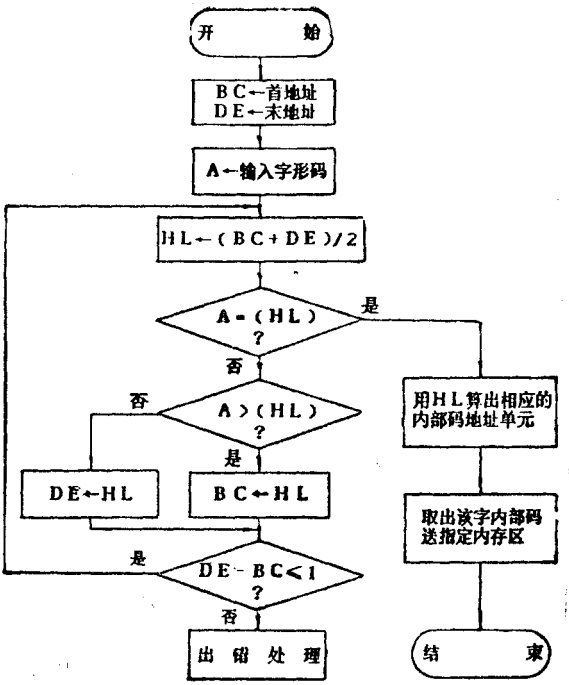


图 6·19 对分查表法流程图

的内存区末地址与首地址之差小于等于 1 时，若仍然没有找到 A = (HL) 的项，就表示在此对照表中没有该输入字形码，于是转入出错处理，告诉操作者不能输入此编码，需查对以后另行输入。

可以证明，对具有n 项的有序对照表来说，对分查表法的平均查找次数为$N = \log_2 n - 1$ 次，是平均查找速度最快的一种线性查表方法。

2. 汉字输出程序

汉字输出程序的主要功能是把经过处理的机器内部码转换成适合于汉字输出设备的数据格式，它是汉字终端应用软件的重要组成部分。

汉字输出设备的种类繁多，汉字输出程序的设计也随不同的输出设备而异。这里主要讨论汉字字形输出程序的基本流程，以及与此密切相关的字形发生器管理程序、显示和打印程序的基本流程。

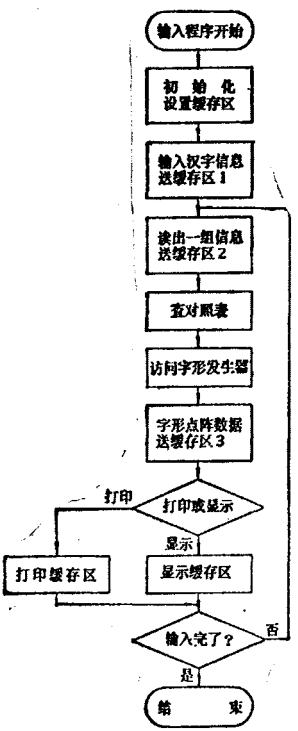


图 6·20 汉字输出的基本流程图

汉字输出的基本流程图如图6·20所示。

如图所示，在程序初始化时，需要设置各种不同用途的数据暂存区，或称缓冲存储区。这里假定设置3个缓存区：缓存区1是存放需要输出的汉字信息内部码，例如文件的一页或一段。缓存区2是存放当前一次输出的汉字内部码，例如一行或一组汉字信息，CPU把它们从缓存区1中读出放入缓存区2中，以便于查表操作及访问字形发生器。缓存区3用于存放字形发生器输出的点阵字形数据，数据的排列格式应根据不同输出设备的硬件结构来决定。打印或

显示设备就是从缓存区3中读取点阵数据、输出汉字字形的。程序流程的最后一项是循环控制，它通常根据在初始化部分设置的一次输出汉字字数，用一个计数器来进行控制，每当输出一个汉字字形数据后，计数器指针便自动减1。当判断计数器不为0时，程序就循环回去继续输出；当判断计数器为0时，就使这一次输出过程结束。

下面分别介绍字形发生器的管理程序和显示、打印程序流程图。

（1）字形发生器管理程序

字形发生器的管理程序通常包括汉字输出流程（参见图6·20）中的“查对照表”和“访问字形发生器”两个内容。它的基本功能是把缓存区2中读出的汉字内部码通过查表方法转换成字形发生器的地址码；然后向字形发生器读取相应的点阵字形数据，送入缓存区3中。查对照表的方法及操作过程与键盘输入管理程序中的查表方法相似，就不再重复了。这里介绍访问字形发生器的操作过程。

如前所述，字形发生器与主CPU的连接方式有内存方式、I/O方式和插接兼容（即汉卡）方式三种。现简单说明以内存方式和I/O方式连接的字形发生器管理程序。

① 内存方式

字形发生器以内存方式连接时，它作为主存储器的一部分，通常是一组固化了字形点阵数据的MASK ROM或EPROM芯片。芯片中的点阵字形数据通常是以字节为单位存放的，一个汉字的16×16点阵字形有32个字节的数据，需占用32个内存地址单元。图6·21是“部、达、到、我”4个汉字的16×16点阵数据在字形发生器中的存放情况以及它们的对照表。

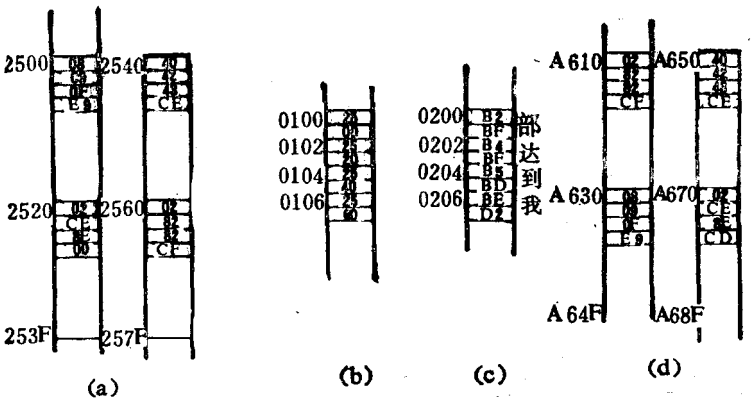


图 6·21　字形发生器及对照表的数据存放格式

图(a)左侧一列数字2500、2520、……2560是4个汉字点阵数据存放的首地址。4个字占用内存地址从2500～257F共128个地址单元。

图(b)、(c)是对照表的两列，分别存放首地址和汉字内部码，每个汉字各占2个字节。由图可知，同一汉字的首地址存放单元与内部码存放单元的地址数是一一对应的。在该图中，两者相差100。于是可以编制出从字形发生器读取字形点阵数据的流程图，如图6·22所示。

首先从图6·21(c)中查出汉字（例如“到”）的内部码(B5BD)的存储地址为0204（注意：图(c)是以内部码第一字节数据递增顺序排列的，故可用对分查表法查出）。将图(c)的地址0204减去固定常数100，得到图(b)中存储汉字“到”的点阵字形数据首地址为0104。于是便可在0104单元和下一个单元0105中读出点阵字形数据首地址2540。CPU从字形发生器的2540地址开始连续读取32个字节数据送到数据缓冲存储区，

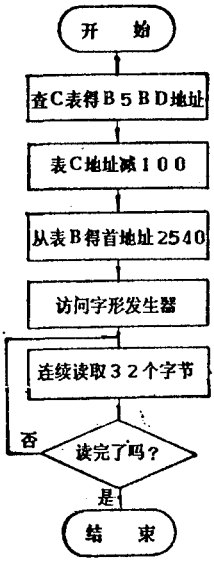


图 6·22 内存方式字形数据输出流程

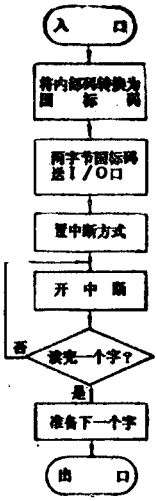


图 6·23 I/O方式字形发生器读取字形数据流程图

见图6·21(d)。输出缓存区中的数据排列格式是按照汉字的读出次序排列存放的。

② I/O方式

在I/O方式连接时，字形发生器是作为CPU的外存储器，它可以是固化数据的MASK ROM芯片，也可以是磁盘驱动器。与内存方式相比，I/O方式的字形发生器有较大的汉字容量，而且不占用CPU的内存资源，故在实际工作中应用十分广泛。

I/O方式字形发生器读取字形数据的流程图如图6·23所示。

在I/O方式时，CPU向字形发生器送入双字节的汉字国标

码就可以读取字形数据。因此，程序的第一步是把汉字内部码转换成国标码，送入字形发生器所在的I/O端口号（即外设口的地址号），然后置中断方式。字形发生器按照输入的国标码两个字节查找存放该汉字字形点阵数据的首地址，找到后向CPU发出中断请求，这时CPU执行一个中断子程序，读取字形发生器送出的该汉字字形点阵数据。

一般说来，I/O方式访问字形发生器的时间与内存方式相比要稍长些，在使用磁盘存储器作为字库时，为了克服重复启动使等待时间过长的缺点，通常在内存中设置一个“在用汉字管理表”，对已经读出存放在字形点阵缓存区的汉字进行登记管理，以后遇到需读取相同汉字时，就可在该缓存区内复制，不必再访问外设了。例如，有1000个字的汉字文件需显示输出，其中有不同的汉字200个，有800个是重复的字。则采用“在用汉字管理表”的办法以后，只需访问字形发生器200次，可以显著加快字形的输出速度。

（2）显示器输出程序

显示器是最常用的汉字字形输出设备。显示输出程序的主要功能是将已存放在输出数据缓冲存储区中的字形数据转换成汉字显示器所需的数据格式，并送入显示器的刷新存储器中，以达到显示汉字字形的目的。

如前所述，字形发生器管理程序输出的字形点阵数据是以汉字的读取次序为顺序排列的（参见图6·21(d)）。而显示缓存区（即刷新存储器）中的数据排列格式要求与屏幕显示点阵格式一一对应，要实现这两种数据格式的转换，必须根据显示器的硬件结构及汉字字形点阵结构的具体情况进行安排。

设有显示器可显示一屏汉字M×N个，其中M是每行汉字

字数，N是显示的汉字行数。被显示的汉字字形为24×24点阵，点阵数据通常是以横排格式从字形发生器读出，存放在输出缓存区中的。图6·24(a)是输出缓存区的数据排列，每个汉字有24个点阵行，每行有3个字节。图中第1个汉字的字形数据为1～72字节，第2个汉字的字形数据为73～144字节，依此类推，第j个汉字的字形数据为72(j－1)＋1～72 j字节。

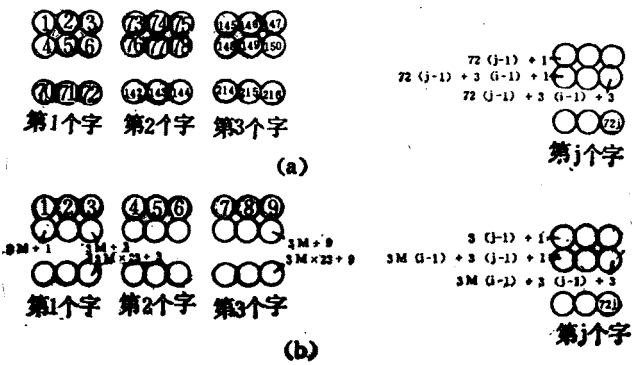


图 6·24　输出缓冲区和显示缓冲区的数据排列

图6·24(b)是显示缓存区的数据排列。如前所述，显示输出时是对一行汉字以点阵行扫描方式同时输出，而不是逐个汉字输出的，因此必须把数据字节排列顺序重新编排。图中假设一行显示M个汉字，则第一个汉字的字形数据是这样组成的：第1点阵行是1～3字节，第2点阵行是3M＋1～3M＋3字节，依此类推，第24点阵行是3M×23＋1～3M×23＋3字节。于是可以算出第j个汉字的第1点阵行是3(j－1)＋1～3(j－1)＋3字节，第2点阵行以下是3M(i－1)＋3(j－1)

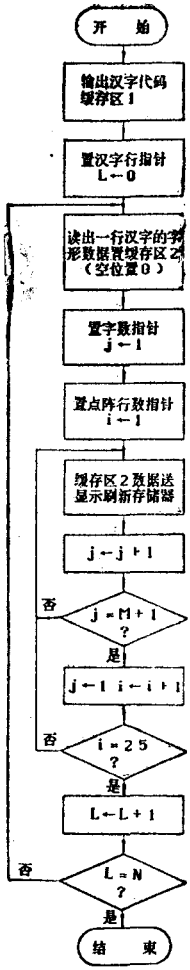


图 6·25 汉字显示输出程序流程图

+ 1～3M(i－1)＋3(j－1)＋3 字节。其中i是点阵行的行数。

于是，我们可以编制显示输出程序流程如图6·25所示。

图中的缓存区2就是图6·24中的输出缓存区，显示缓存区就是显示刷新存储器。从“开始”以下的第6个方框：“缓存区2数据送显示刷新存储器”，它们之间的字节数换算关系可以归纳如下算法（适用于24×24点阵字形）：

取缓存区2第j个字的字节数	送显示刷新存储器的字节数
3(i-1)+1	72·L·M+3(j-1)+3(i-1)M+1
3(i-1)+2	72·L·M+3(j-1)+3(i-1)M+2
3(i-1)+3	72·L·M+3(j-1)+3(i-1)M+3

在上式中，L是汉字行数，

M是每行汉字数，

i是每字的点阵行数。

框图中N是显示一屏汉字的最大行数，当L=N时，显示数据的格式转换过程结束。

（3）打印输出程序

打印输出程序的主要功能是把输出缓存区中的字形点阵数据转换成适合打印设备特点的数据格式，并送入打印缓存区中。

由针式打印机的原理可知，它是按照终端提供的字形数据信号，用直排1列或2列钢针横向（水平）移动的方法，逐列打印汉字的各列点阵，形成汉字字形输出的。因此，它与显示器输出的情况不同，构成字形的点阵数据应是直排格式的（参见图4·37）。打印输出程序必须把字形发生器输出的横排格式字形数据转换成直排格式，并根据打印设备硬件结构的具体要求存放在打印缓存区。在打印输出过程中，输出程序还要不断测

试打印设备的工作状态，协调主机与打印机在时间上的差异。

打印输出程序的流程图如图6·26所示。首先，终端把需打印输出的汉字内部码存入输出缓存区（缓存区1），然后根据打印行宽，读取一行汉字的字形点阵数据存入缓存区2（点阵缓存区）。置打印缓存区参数指针以后，将缓存区2内的数据按照打印机所需的格式存入打印缓存区，存入完了以后，再置打印机的各项参数指针，启动打印机进行打印输出。当一行汉字打印完后，检测缓存区1中汉字是否输出完了？若未完，重新开始下一行打印；否则，打印输出程序结束。

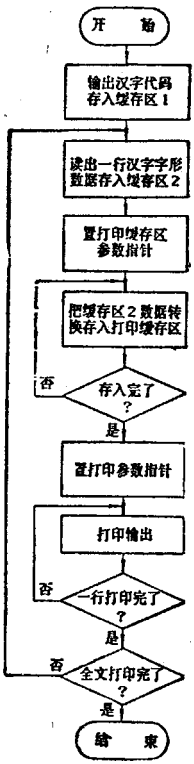


图 6·26 打印输出程序流程图

3. 汉字文本编辑程序

汉字文本编辑程序的主要功能是对显示在屏幕上的汉字文本进行修改补充，最后以指定的格式输出（显示或打印）。

文本编辑涉及的内容很多，这里主要介绍汉字单字和行的增、删、改等简单编辑功能的程序流程。

（1）光标定位程序

光标是显示屏幕物理位置的指示标记。光标定位程序的作

用是将光标移动到编辑文件所需要的指示位置上，以便进行文字的删改和插入。

为了进行光标定位，我们把显示屏幕作为一个二维平面矩阵D，每个汉字在屏幕上的位置可以用矩阵元素d_{ij}表示，其中$1 \leqslant i \leqslant n$，$1 \leqslant j \leqslant m$。这里，n是屏幕能显示的最大汉字行数，m是每行的最大字数。这样，光标右移就是矩阵元素d_{ij}在当前位置上的列数j增加1，光标左移就是j减少1；而光标下移就是矩阵元素d_{ij}在当前位置上的列数不变，行数i增加1，光标上移就是i减少1。当j等于m时，若要光标右移，则应把光标置在下一行的第一个字位上（i增加1，j=1）。依此类推，我们可以画出光标定位程序的流程图如图6·27所示。

图中，在入口开始后，输入光标定位代码K。主机通过检测K的不同值，确定执行哪一种光标定位操作。当K=0时，置i=1，j=1，就是把光标移到屏幕左上角的起点位置。

当i=1时，若执行光标上移操作，就需进行溢出处理（还有其它类似三种情况、见图6·27）。溢出处理的常用方法是转入屏幕滚动（下滚或上滚）子程序，使光标重新进入屏幕矩阵D中。

如前所述，主机CPU对汉字信息的加工处理是以内部码的形式进行的，一个汉字的内部码通常是两个字节，它在屏幕上所占的物理位置也是一个西文字符的两倍。因此，可以在内存中设置一个与屏幕矩阵D相对应的汉字代码映照区，也就是以汉字内部码表示的显示缓存区，使映照区内的汉字代码所在的地址单元，与屏幕显示汉字所在矩阵元素一一对应。这样，可以把光标移动转换成映照区（缓存区）地址指针的移动，实现对文件内容的编辑。

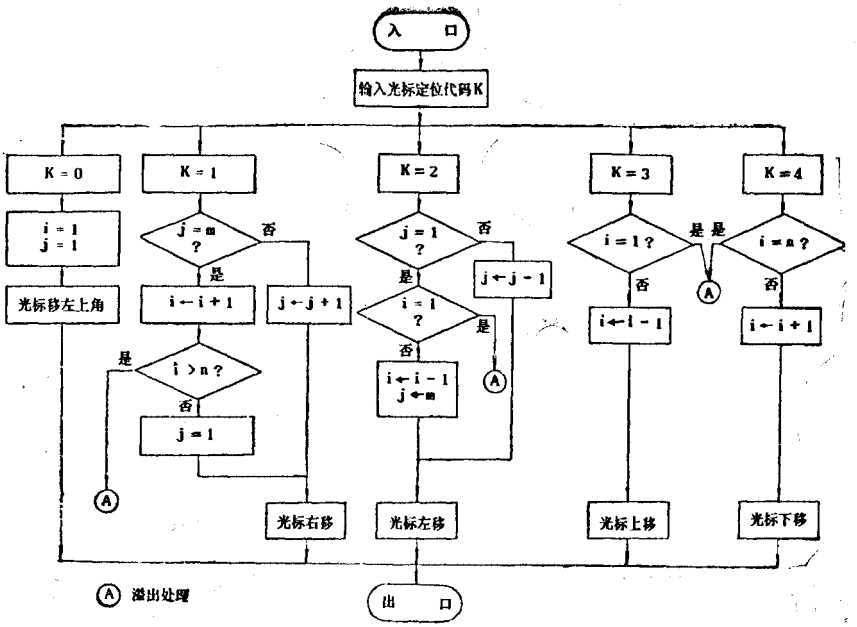


图 6·27 光标定位程序流程图

(2) 汉字的增、删、改程序

对汉字的增、删、改是文件编辑中最常用的基本操作，可以调用汉字增、删、改程序来完成。

① 增加汉字

实现增加汉字的方法是先用光标定位程序确定需增加汉字在文件中的位置，然后依次将指定位置以后的汉字代码向后移动，空出光标指示的位置。便可将增加汉字的代码置入。增加汉字的程序流程如图6·28所示。

如果显示已达最后位置：K=n，L=m，（n、m是显示

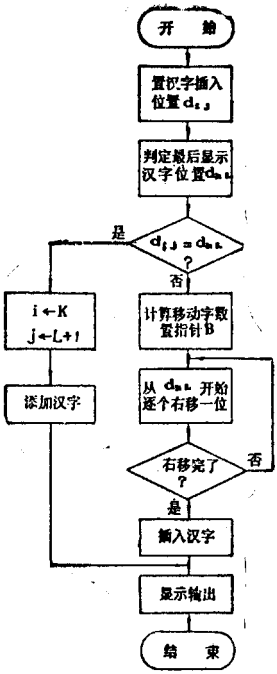


图 6·28 增加汉字程序流程图

屏幕的最大显示行数和每行最大字数）则必须先把最后一行内容转送到一个缓存区；尔后执行上述程序流程。最后再将缓存区内容添加在修改后的文件的末尾。

② 删除汉字

实现删除汉字的方法与增加汉字相似，先用光标定位程序确定删除汉字的位置d_{ij}，然后从$d_{i,j+1}$的矩阵位置开始，依次执行左移一位程序，直到最后的汉字位置$d_{K,L}$移到$d_{K,L-1}$，并在$d_{K,L}$位置上填入一个空白字符，最后把删改好的文件显示输出。

③更改汉字

更改汉字的程序流程比较简单。先用光标定位程序指出被更改的汉字所在位置，然后将新的汉字内部码写入这个位置，原来有错的字就自动被抹掉，最后进行显示输出。

如果更改的字数与原来有错的字数不相等，则应先执行插入或删除程序，使两者的字数相同，然后执行更改程序。

(3) 汉字行的增加或删除程序

在汉字文件编辑过程中，需要增加或删除一行或几行汉字时，可以使用汉字行的增加或删除程序。

① 汉字行的增加

若需增加一行汉字时，应先检查原文件中是否有可压缩的空白行。若空白行在正文与标题之间，则把正文第一行开始逐行上移一行；如果空白行在正文之后，则把正文最后一行开始逐行下移一行。然后可将正文中空出的一行插入需增加的内容。若没有空白行，则需将正文的最后一行内容移入下一页文件的第一行，再依次执行下移一行的操作。增加一行的程序流程图如图6·29所示。

设在正文的第l+1行插入新的一行内容，置正文行数指针L_0，若正文与标题间有空行，则把首行行数送入L_0，把l行以上各行逐一上移；若空行在行末，则把末行行数送入L_0，把l+1行以下各行逐一下移。完成上移或下移操作之后，把新增的内容插入l+1行，程序结束。图中$L_0-1\leftarrow(L_0)$是表示把L_0所指一行的内容送入上一行中。

② 汉字行的删除

当需要删除文件中的某一汉字行时，先置入删除行的行号，然后从给出的行号开始，将该行以后的每一行内容顺序地

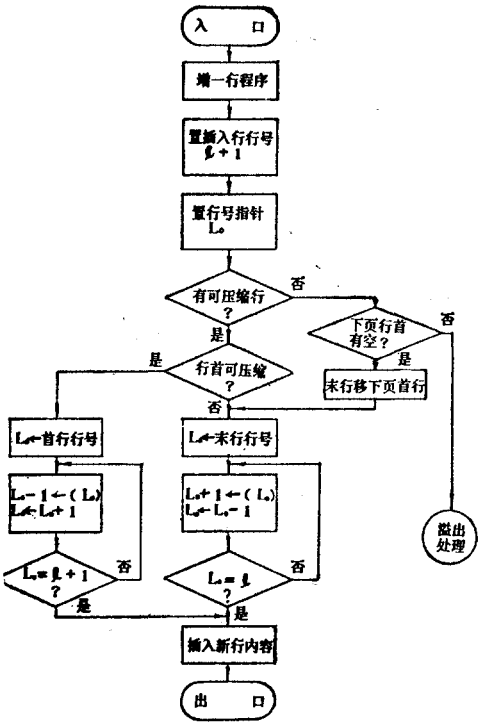


图 6·29 增加一行的程序流程图

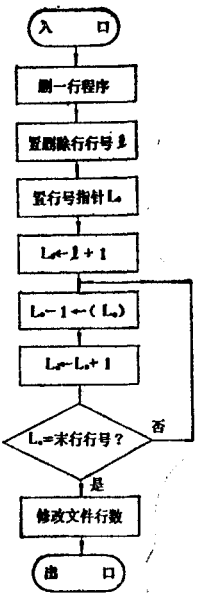


图 6·30 删除一行的程序流程图

填入前一行，直至文件的最后一行上移一行为止。删除一行的程序流程如图6·30所示。

图中没有单独做一次第l行内容的删除操作，因为在执行把第l+1行内容上移一行的操作时（图中第6框：$L_0-1\leftarrow L_0$)），就自动删除了第l行的原有内容了。但在上述操作之前，必须把每行中的空白处全部置“0”。在删除完毕后应修改文件的总行数，然后程序结束。

主要参考资料

1. 汉字信息处理技术　郭平欣张淞芝主编　国防工业出版社
2. 中文信息的计算机处理　张寿萱等著　宇航出版社
3. 计算机的中文信息处理（上册）　赵珀璋等编著　宇航出版社　87.12
4. 中文信息处理系统　陈增武金连甫编著《计算机应用通信》
5. 中文信息处理国际研讨会论文集 1～3 集　1983 12.
6. 国家标准《信息交换用汉字编码字符集》GB2312.
7. 国家标准《信息处理交换用七位编码字符集》GB1988
8. 国家标准《信息处理交换用七位编码字符集的扩充方法》GB2311
9. 国家标准《汉字整字键盘盘面字排列》GB7513
10. 国家标准《数据终端设备和数据电路终接设备之间的接口电路定义表》GB3454
11. 国家标准《数据通信基本型控制规程》GB3453
12. 电子电传机　薛华日编著　人民邮电出版社
13. 数字通信　毕光国编著　人民邮电出版社
14. 数据通信手册　海文学等译　人民邮电出版社
15. 显示技术与显示器件　彭国贤编著　人民邮电出版社
16. 图象识别技术及其应用　李月景编著　机械工业出版社
17. 中国大百科全书　电子学与计算机卷（Ⅰ、Ⅱ）
18. 中国大百科全书　交通卷
19. 微型计算机　朱绍庐主编　人民交通出版社
20. 微型计算机原理及其应用　王长胤等编　湖北科技出版社
21. 微型计算机原理与应用　朱仲英编　上海交大出版社
22. Z80汇编语言程序设计　重庆科技文献出版社
23. 中文信息（双月刊）　1984～1988.
24. 各种汉字输入编码方案的资料、手册.
25. 汉字信息处理　邱光谊、周有文等编著　人民邮电出版社　88.3

电信技术普及丛书

人民邮电出版社出版　各地新华书店发行

现已出版28种，书目如下：

近代电信系统简说	周炯槃、钟义信编著
无线电波传播	程新民编著
微波接力通信	焦其祥　籍义忠编著
卫星通信（修订本）	王德文编著
脉码调制通信（修订本）	《脉码调制通信》编写组
数据通信	赵　辰　彭美云编著
毫米波波导通信	邮电部武汉邮电科学研究院编写组
激光通信	邮电部武汉邮电科学研究院编写组
散射通信	李道本编著
纵横制电话交换机浅说	朱景惠编
传真通信	邮电532厂编写组
图象通信	余兆明编著
用户电报	高星忠　朱筱英编著
程控电报信息交换	杨晋儒编著
电子电传机	薛华日编著
中文电报译码机（修订本）	张　理编著
汉字终端技术入门	周关兴编著
半导体集成电路	邮电508厂编写组
模拟集成电路	赵保经编著
电视电话	余兆明等编著
彩色电视（修订本）	肖克昌编著

大屏幕电视	本书编写组
磁带录象技术	张夏阳　赵力飒编著
模式识别	陈尚勤　魏鸿俊编著
显示技术与显示器件	彭国贤编著
速调管	邮电506厂《速调管》编写组
行波管	邮电506厂《行波管》编写组
移动通信	张以进编著

果粉藏書

科技新书目：212-122
ISBN 7-115-04161-X/TN·325
定　价：5.40 元